



सत्यमेव जयते

INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI

I.A.R.I.6.

GIP NLK—H 3 I.A.R.I. - 10-5 55- 15,000

B The Biometrics Section of the American Statistical Association I O M E T R I C S

TABLE OF CONTENTS

An Open Letter to Biometrics' Subscribers

Triple Rectangular Lattices BOYD HARSHBARGER 1

Problems of the Optimum Catch in Small Whitefish Lakes
RICHARD B. MILLER 14

Statistics of a Lake Trout Fishery F. E. J. FRY 27

Queries 68

Abstracts 76

The Biometric Society 90

News and Notes 94



48039

Number 1	March 1949	Volume 5
----------	------------	----------

Tindlithgow Library.
 Imperial Agricultural Research Institute
 New Delhi.

AN OPEN LETTER TO BIOMETRICS' SUBSCRIBERS

BECAUSE OF THE current rise in costs, the Board of the American Statistical Association has found it necessary to restrict the number of pages of its journals for the coming year in order to keep expenses within the amount of funds currently available. For *Biometrics*, a limit of 64 pages per issue has been recommended by the Board subject to final approval by the Council plus any additional pages purchased by the Biometric Society.

Before this decision was reached the editors of *Biometrics* had planned on approximately 80 pages for each of the four issues during 1949. Contributors have supplied enough valuable papers to make this easily possible. When the above decision by the Board became known, a number of individual subscribers to *Biometrics*, feeling that reduction of the size of the journal at this time would be undesirable, banded together informally in an effort to permit continued growth in the size of *Biometrics*. The Association quoted a price of \$20.00 a page to this group, this being the price charged the Biometric Society.

Funds contributed immediately were sufficient to increase the present issue to 98 pages. Funds for additional pages in the next three issues are solicited. If you agree that the publication of valuable papers such as have been appearing in *Biometrics* should be facilitated rather than restricted, please send your contribution at once to the editor:

Miss Gertrude Cox, Editor
c/o Institute of Statistics
State College Station
Raleigh, North Carolina

In this regard, the following information may be of interest. It has been suggested by the Board of the American Statistical Association that the Institute of Statistics, Inc. at North Carolina State College, Raleigh, North Carolina be requested to publish *Biometrics*. Negotiations are in progress between the Association and the Institute to make this change following publication of the present (March 1949) issue. If this move is completed, the cost for additional pages may be somewhat lower than the \$20.00 quoted above. In any case the price will not be more than \$20.00 a page, and all money contributed for this purpose will be used to expand *Biometrics*.

For "*Friends of Biometrics*"

J. Berkson
M. C. Bruyere
J. W. Tukey
C. P. Winsor

TRIPLE RECTANGULAR LATTICES

BOYD HARSHBARGER

Virginia Agricultural Experiment Station

THE THEORY OF the experimental arrangement which is now called the lattice as given by Cochran [1] and Yates [2,3,4] requires that the number of varieties be an exact square. To avoid the restriction that the number of varieties or treatments be a perfect square, Yates [2] introduced the design which he called pseudo-factorial with unequal groups of sets. No attempt was made to use the recovery of information, and the design proposed by him does not conveniently lend itself to such an analysis.

This paper and a previous one, *Rectangular Lattices*, [5] by the author, present an extension of the incomplete block designs to the cases where the number of varieties or treatments are expressible as the product of two integers. An explicit solution is given for the cases where the number of varieties are two consecutive integers. This paper treats, in particular, the cases where there are three groups of sets. These arrangements differ from Yates' non-square design since the blocks are all the same size and the varieties are adjusted by both the inter- and intra-block information. The name triple rectangular lattices is proposed for the design where there are three groups of sets.

As in the square triple lattices, the varieties for the triple rectangular lattices are arranged in three groups, X, Y, and Z, each of which is replicated as shown below. For simplicity of illustration, a 4×3 triple rectangular lattice is used. The numbers designate varieties.

In practice the varieties are randomized within blocks and the blocks are randomized within the replicates.

For purposes of enumeration and computation triple rectangular lattices may be thought of as square lattices with k varieties missing in the main diagonal. With this arrangement the groups for a 4×3 triple rectangular lattice (together with more convenient subscripts) are shown in Table II.

Group Y is formed from group X by using the columns as blocks; group Z is formed by superimposing the Latin square of Table III upon group X.

TABLE I

GROUP X							
Blocks				Blocks			
(1)	1	2	3	(1)	1	2	3
(2)	4	5	6	(2)	4	5	6
(3)	7	8	9	(3)	7	8	9
(4)	10	11	12	(4)	10	11	12

GROUP Y							
Blocks				Blocks			
(1)	4	7	10	(1)	4	7	10
(2)	1	8	11	(2)	1	8	11
(3)	2	5	12	(3)	2	5	12
(4)	3	6	9	(4)	3	6	9

GROUP Z							
Blocks				Blocks			
(1)	5	9	11	(1)	5	9	11
(2)	3	7	12	(2)	3	7	12
(3)	1	6	10	(3)	1	6	10
(4)	2	4	8	(4)	2	4	8

Table II serves a double purpose. The body of the table gives the pattern of the arrangement in the blocks of the $k(k-1)$ varieties, the V 's being simply symbols for varieties. However, if each V is regarded as the total of the two observations on a variety which runs in the two replicates of the group, then this table is one which is made up in the course of the analysis.

The use of the Latin squares for getting the arrangement in the third group is perfectly general except for squares which have the property of the 6 by 6 and 10 by 10 Latin square. The solution of the lattice is simplified if the columns are interchanged in the Latin square so as to put the A, B, C —in order along the main diagonal.

The mathematics for these statistical designs is similar to that used in the development of the rectangular lattices. Following the method used in the author's publication, *Rectangular Lattices*, the necessary formulas and equations are evolved for the analysis of the triple rectangular lattices. The symbols used in the paper are defined as follows:

1. R_{hj} is the sum of the yields of the varieties for replicate j of group h .

TABLE II

GROUP X				
Blocks				
(1)	0	V_{123}	V_{134}	V_{142}
(2)	V_{214}	0	V_{231}	V_{243}
(3)	V_{312}	V_{324}	0	V_{341}
(4)	V_{413}	V_{421}	V_{432}	0

GROUP Y				
Blocks				
(1)	0	V_{214}	V_{312}	V_{413}
(2)	V_{123}	0	V_{324}	V_{421}
(3)	V_{134}	V_{231}	0	V_{432}
(4)	V_{142}	V_{243}	V_{341}	0

GROUP Z				
Blocks				
(1)	0	V_{231}	V_{341}	V_{421}
(2)	V_{142}	0	V_{312}	V_{432}
(3)	V_{123}	V_{243}	0	V_{413}
(4)	V_{134}	V_{214}	V_{324}	0

TABLE III

A	C	D	B
D	B	A	C
B	D	C	A
C	A	B	D

2. A_{hi} is the difference between the yields of the varieties in blocks i of group h .
3. B_{hi} is the total of the yields of the varieties from the blocks i of group h .
4. T_{hi} is the sum of the yields from all six replicates of the varieties listed in block i of group h .
5. V_{efg} is the sum of the yields from all six replicates of the variety with subscript efg .
6. G is the grand total of the yields of all the varieties in the six replicates.
7. y_{efg} is the yield of a variety with subscript efg for a particular replicate and a particular group.
8. k is the number of blocks in a replicate.

Analysis of Variance Table		
Source of Variation	Degrees of Freedom	Sum of Squares
Replicates	5	$\sum_h^{x,y,z} \frac{R_{hi}^2}{k(k-1)} - \frac{G^2}{6k(k-1)}$ (3)
Component (a)	$3(k-1)$	$\sum_h^{x,y,z} \frac{A_{hi}^2}{2(k-1)} - \frac{\sum_h^{x,y,z} (R_{hi} - R_{h2})^2}{2k(k-1)}$ (4)
Component (b)	$3(k-1)$	$\frac{1}{12k(2k-3)} \left\{ \sum_h^{x,y,z} (2k-1)(3B_{hi} - T_{hi})^2 \right.$ (5)
		$\left. - 2 \sum_{i=1}^k [(3B_{xi} - T_{xi})(3B_{yi} - T_{yi}) + (3B_{xi} - T_{xi})(3B_{zi} - T_{zi}) \right.$ (6)
		$\left. + (3B_{yi} - T_{yi})(3B_{zi} - T_{zi}) \right] - 2 \sum_h^{x,y,z} [3(R_{hi} + R_{h2}) - G]^2 \}$
Varieties (unadjusted)	$k(k-1) - 1$	$\sum V_{x/2}^2 - \frac{G^2}{6k(k-1)}$ (7)
Error (intra-block)	$5k^2 - 11k + 1$	by subtraction
Total	$6k(k-1) - 1$	$Sy_{x/2}^2 - \frac{G^2}{6k(k-1)}$ (8)

In the method of analysis used, the variety means are adjusted by using both the inter- and intra-block variance. This necessitates finding two weights W and W' . These weights are also used in calculating the standard errors and the efficiency of the design.

The weights are calculated by two simple formulas

$$\frac{1}{W} = E \quad (1)$$

and

$$\frac{1}{W'} = \frac{6B - E}{5} \quad (2)$$

where B is the average mean square for component (a) and (b),

E is the mean square for the error (intra-block).

The analysis of variance table for this design is given in Table IV.

The adjusting of the varietal means using both the inter- and the intra-block information is accomplished by calculating certain constants. These constants are then subtracted from the varietal means. If the varietal means are arranged in the order of group X and with the Latin letters superimposed, then the constants to be subtracted from row i , column j , and Latin letter k are as follows:

$$c_{xj} = \quad (9)$$

$$\frac{(W - W') \{ [k(2W + W') - 3W'] (3B_{xj} - T_{xj}) - (W - W') \sum_h^{v,z} (3B_{hx} - T_{hx}) \}}{6[k(2W + W') - 3W'] [k(2W + W') - 3W']}$$

$$c_{yj} = \quad (10)$$

$$\frac{(W - W') \{ [k(2W + W') - 3W'] (3B_{yj} - T_{yj}) - (W - W') \sum_h^{v,z} (3B_{hy} - T_{hy}) \}}{6[k(2W + W') - 3W'] [k(2W + W') - 3W']}$$

$$c_{zk} = \quad (11)$$

$$\frac{(W - W') \{ [k(2W + W') - 3W'] (3B_{zk} - T_{zk}) - (W - W') \sum_h^{v,z} (3B_{hz} - T_{hz}) \}}{6[k(2W + W') - 3W'] [k(2W + W') - 3W']}$$

To test the significance of two varieties the variance of their difference is needed. In the triple rectangular lattice designs there are seven formulas for the variances depending upon the combination of the varietal means.

The formulas for the variances of two varieties appearing together in same block with

1 digit alike in the subscripts

as

$$\begin{aligned} \text{Variance } (V_{135} - V_{134}) \\ = \frac{1}{3W} \left\{ 1 + \frac{2(W - W')[k(2W + W') - (W + 2W')]}{[k(2W + W') - 3W][k(2W + W') - 3W']} \right\}; \end{aligned} \quad (12)$$

2 digits alike in the subscripts

as

$$\begin{aligned} \text{Variance } (V_{125} - V_{421}) \\ = \frac{1}{3W} \left\{ 1 + \frac{(W - W')[2k(2W + W') - (W + 5W')]}{[k(2W + W') - 3W][k(2W + W') - 3W']} \right\}; \end{aligned} \quad (13)$$

3 digits alike in the subscripts

as

$$\text{Variance } (V_{125} - V_{152}) = \frac{1}{3W} \left\{ 1 + \frac{2(W - W')}{k(2W + W') - 3W} \right\}. \quad (14)$$

The formulas for the variance of two varieties not appearing in same block

No digits alike in the subscripts

as

$$\begin{aligned} \text{Variance } (V_{126} - V_{754}) \\ = \frac{1}{3W} \left\{ 1 + \frac{3(W - W')[k(2W + W') - (W + 2W')]}{[k(2W + W') - 3W][k(2W + W') - 3W']} \right\}; \end{aligned} \quad (15)$$

1 digit alike in the subscripts

as

$$\begin{aligned} \text{Variance } (V_{125} - V_{342}) \\ = \frac{1}{3W} \left\{ 1 + \frac{(W - W')[3k(2W + W') - (2W + 7W')]}{[k(2W + W') - 3W][k(2W + W') - 3W']} \right\}; \end{aligned} \quad (16)$$

2 digits alike in the subscripts
as

$$\text{Variance } (V_{123} - V_{231}) \\ = \frac{1}{3W} \left\{ 1 + \frac{(W - W')[3k(2W + W') - (W + 8W')]}{[k(2W + W') - 3W][k(2W + W') - 3W']} \right\}; \quad (17)$$

3 digits alike in the subscripts
as

$$\text{Variance } (V_{123} - V_{612}) = \frac{1}{3W} \left[1 + \frac{3(W - W')}{k(2W + W') - 3W} \right]. \quad (18)$$

To illustrate numerically the method of analysis for the triple rectangular lattices the analysis is presented of an alfalfa variety experiment¹ performed in 1948 at the Piedmont Field Station near Orange, Virginia.

There were 12 varieties to be tested in the experiment so a 4×3 rectangular lattice was used. The experiment was set up as in Table I and when placed in the field, the varieties were randomized within each block and the blocks within each replicate.

After the first cutting of the alfalfa, the yields were tabulated and compiled for computational purposes as shown in Table VI. The upper figure in each plot refers to the variety number while the lower figure is the variety yield in pounds per plot (green weight). The size of the plot was 2×20 feet.

In Table VII the yields of each variety have been totaled by groups, that is, yields of the same variety in replicate 1 and 2 of group X were added together, etc.

The total yields of the 12 varieties are given in Table VIII. Here the three group yields for each variety were added together and the rows, columns, and Latin letter totals recorded.

The calculations for the analysis of variance are derived by substituting in formulas (3) through (8).

In order to substitute in formulas (4) and (5) (the block components sums of squares) it is convenient to form Table IX.

The component (a) set of differences are the differences between the sums of yields for the 3 varieties appearing together in block i of replicate

¹The alfalfa data are used through the courtesy of Agronomists T. J. Smith and G. D. Jones of the Virginia Agricultural Experiment Station. Mrs. Sally R. Hudgins of the Virginia Agricultural Experiment Station is responsible for the numerical analysis.

TABLE VI
YIELD OF ALFALFA VARIETIES BY REPLICATIONS

GROUP X													
Replicate 1						Replicate 2							
Blocks						Block Totals	Blocks					Block Totals	
(1)	0	1	2	3			(1)	0	1	2	3		
		13.06	5.68	6.28	25.02				10.70	4.36	5.66	20.72	
(2)	4		5	6			(2)	4		5	6		
	8.24	0	8.32	7.84	24.40			10.34	0	6.44	10.06	26.84	
(3)	7	8		9			(3)	7	8		9		
	7.32	6.86	0	5.04	19.22			5.62	7.90	0	7.70	21.22	
(4)	10	11	12				(4)	10	11	12			
	8.88	11.42	10.38	0	30.68			6.46	8.36	6.74	0	21.56	
Total					(R_{x1})	99.32						(R_{x2})	90.34

GROUP Y													
Replicate 1						Replicate 2							
Blocks						Block Totals	Blocks					Block Totals	
(1)	0	4	7	10			(1)	0	4	7	10		
		8.55	9.72	4.00	22.27				10.74	8.18	8.92	27.84	
(2)	1		8	11			(2)	1		8	11		
	10.56	0	6.60	9.64	26.80			12.62	0	8.52	11.92	33.06	
(3)	2	5		12			(3)	2	5		12		
	6.76	8.60	0	8.06	23.42			9.18	9.76	0	8.34	27.28	
(4)	3	6	9				(4)	3	6	9			
	7.60	7.82	7.98	0	23.40			7.76	10.38	10.70	0	28.84	
Total					(R_{y1})	95.89						(R_{y2})	117.02

GROUP Z

Replicate 1						Replicate 2					
Blocks					Block Totals	Blocks					Block Totals
		5	9	11				5	9	11	
(1)	0	9.86	9.28	13.04	32.18	(1)	0	8.68	9.40	9.98	28.06
	3		7	12			3		7	12	
(2)	8.74	0	9.34	10.68	28.76	(2)	5.46	0	9.41	10.52	25.39
	1	6		10			1	6		10	
(3)	11.36	8.52	0	6.32	26.20	(3)	14.02	11.76	0	8.84	34.62
	2	4	8				2	4	8		
(4)	5.54	10.58	8.88	0	25.00	(4)	8.96	12.00	9.64	0	30.60
Total					(R_{z1}) 112.14						(R_{z2}) 118.67

TABLE VII
VARIETY YIELDS BY GROUPS

GROUP X					
				B_{xi}	$3(B_{xi})$
	1	2	3		
0	23.76	10.04	11.94	45.74 (B_{x1})	137.22
4		5	6		
18.58	0	14.76	17.90	51.24 (B_{x2})	153.72
7	8		9		
12.94	14.76	0	12.74	40.44 (B_{x3})	121.32
10	11	12			
15.34	19.78	17.12	0	52.24 (B_{x4})	156.72
Group totals				189.66	568.98

TABLE VII—Continued

GROUP Y					
				B_{y_i}	$3(B_{y_i})$
	4	7	10		
0	19.29	17.90	12.92	50.11 (B_{y1})	150.33
1		8	11		
23.18	0	15.12	21.56	59.86 (B_{y2})	179.58
2	5		12		
15.94	18.36	0	16.40	50.70 (B_{y3})	152.10
3	6	9			
15.36	18.20	18.68	0	52.24 (B_{y4})	156.72
Group totals				212.91	638.73

GROUP Z					
				B_{z_i}	$3(B_{z_i})$
	5	9	11		
0	18.54	18.68	23.02	60.24 (B_{z1})	180.72
3		7	12		
14.20	0	18.75	21.20	54.15 (B_{z2})	162.45
1	6		10		
25.38	20.28	0	15.16	60.82 (B_{z3})	182.46
2	4	8			
14.50	22.58	18.52	0	55.60 (B_{z4})	166.80
Group totals				230.81	692.43

1 of a group h and the sums of the yields of the same varieties appearing in block i of replicate 2 of the same group. The component (b) set of differences is obtained by subtracting the T_{hi} values of Table VIII from the $3 B_{hi}$ values of Table VII. The two sets of differences are given in Table IX.

TABLE VIII
VARIETY TOTAL YIELDS

				Row Totals	Latin Letter Totals
	1	2	3		
0	72.32	40.48	41.50	154.30 (T_{21})	A 166.12 (T_{21})
4		5	6		
60.45	0	51.66	56.38	168.49 (T_{22})	B 145.81 (T_{22})
7	8		9		
49.59	48.40	0	50.10	148.09 (T_{23})	C 172.12 (T_{23})
10	11	12			
43.42	64.36	54.72	0	162.50 (T_{24})	D 149.33 (T_{24})
Column Totals					
153.46 (T_{11})	185.08 (T_{12})	146.86 (T_{13})	147.98 (T_{14})	633.38 (G)	633.38

The results of the analysis of variance for the experiment are shown in Table X.

In adjusting the variety means using the inter- and intra-block variance, the weights and correction terms are calculated from formulas (1) and (2) and formulas (9), (10), and (11), respectively. W was found to be .75632 and W' to be .35427. The unadjusted variety means and the correction terms (c_{21} , c_{22} , and c_{23} values) are shown in Table XI. Subtracting the appropriate corrections from each unadjusted variety mean gives the adjusted variety means shown in Table XII.

The standard error of the difference between the means of two varieties occurring together in the same block and

- (1) having one digit alike in the subscripts is .710;
- (2) having two digits alike in the subscripts is .712;
- (3) having three digits alike in the subscripts is .713.

The standard error for two varieties not occurring together in the same block and

- (1) having one digit alike in the subscripts is .734;
- (2) having two digits alike in the subscripts is .735.

TABLE IX
COMPONENT (a) BLOCK DIFFERENCES

Group X	Group Y	Group Z
(A _{x1}) 4 30	(A _{y1}) -5 57	(A _{z1}) 4 12
(A _{x2}) -2 44	(A _{y2}) -6 26	(A _{z2}) 3 37
(A _{x3}) -2 00	(A _{y3}) -3 86	(A _{z3}) -8 12
(A _{x4}) 9 12	(A _{y4}) -5 44	(A _{z4}) -5 60
(R _{x1} - R _{x2}) 8 98	(R _{y1} - R _{y2}) -21 13	(R _{z1} - R _{z2}) -6 53

COMPONENT (b) BLOCK DIFFERENCES

$3B_{x1} - T_{x1}$	$3B_{y1} - T_{y1}$	$3B_{z1} - T_{z1}$
(x ₁) -17 08	(y ₁) -3 13	(z ₁) 14 60
(x ₂) -14 77	(y ₂) -5 50	(z ₂) 16 64
(x ₃) -26 77	(y ₃) 5 24	(z ₃) 10 31
(x ₄) -5 78	(y ₄) 8 74	(z ₄) 17 17
Totals -64 40	5 35	59 05

TABLE X
THE ANALYSIS OF VARIANCE OF A TRIPLE RECTANGULAR LATTICE EXPERIMENT

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Replicates	5	59 216	11 843
Component (a)	9	35 745	3 972
Component (b)	9	10 562	1 171
Blocks (eliminating varieties)	18	46 307	2 5720(B)
Varieties (unadjusted)	11	165 809	15 082
Error	37	48 920	1 3222(B)
Total	71	320 342	

REFERENCES

- [1] G. M. Cox, R. C. Eckhart and W. G. Cochran, "The Analysis of Lattice and Triple Lattice Experiments in Corn Varietal Tests", *Iowa Agr. Exp. Sta. Res. Bul.*, Vol. 281 (1940).
- [2] F. Yates, "A New Method of Arranging Variety Trials Involving a Large Number of Varieties", *Journal Agr. Sci.*, Vol. 26 (1936) pp. 424-455

TABLE XI
UNADJUSTED VARIETY MEAN YIELDS

				c_{zi}	c_{zi}
	1 (C)	2 (D)	3 (B)		
0	12.053	6.745	6.916	$(c_{z1}) - .216$	$(A) .193 (c_{z1})$
4 (D)		5 (A)	6 (C)		
10.075	0	8.610	9.396	$(c_{z2}) - .187$	$(B) .217 (c_{z2})$
7 (B)	8 (D)		9 (A)		
8.265	8.066	0	8.350	$(c_{z3}) - .336$	$(C) .134 (c_{z3})$
10 (C)	11 (A)	12 (B)			
7.236	10.726	9.120	0	$(c_{z4}) - .091$	$(D) .209 (c_{z4})$
c_{yi}	(c_{y1})	(c_{y2})	(c_{y3})	(c_{y4})	
	-.036	-.068	.077	.096	

TABLE XII
ADJUSTED VARIETY MEAN YIELDS

Variety	Mean	Variety	Mean
1.	12.20	7.	8.42
2.	6.68	8.	8.26
3.	6.82	9.	8.40
4.	10.09	10.	7.23
5.	8.53	11.	10.69
6.	9.35	12.	8.92

REFERENCES CONTINUED

- [3] F. Yates, "The Recovery of Inter-Block Information in Three Dimensional Lattice", *Annals of Eugenics*, Vol. 9 (1939) pp. 136-156.
- [4] F. Yates, "The Recovery of Inter-Block Information in Balanced Incomplete Block Designs", *Annals of Eugenics*, Vol. 10 (1940) pp. 317-325.
- [5] Harshbarger, Boyd, "Rectangular Lattices", *Virginia Agricultural Experiment Station, Memoir 1*, 1947.

PROBLEMS OF THE OPTIMUM CATCH IN SMALL WHITEFISH LAKES

RICHARD B. MILLER

THE WHITEFISH fishery of the Canadian prairies is pursued in hundreds of widely scattered lakes, most of which are small and shallow, but highly productive. The whitefish, *Coregonus clupeaformis* Mitchill, thrives in them, despite the fact that they warm to the bottom every summer and produce heavy algal blooms. In the province of Alberta, these lakes range in size from ten to over four hundred square miles in area and each produces from fifty thousand to over half a million pounds of whitefish per year. The annual catch of whitefish from these lakes is governed by the quota system; each lake has a limit set for each fishing season; when the limit has been taken, the local fish inspector closes the lake. If, at the time the lake is closed, the fishing is still good and the season not advanced, it is common practice to grant extensions and re-open lakes for a further limit.

From time to time, one or another of these lakes fails to produce whitefish of commercial size when the fishing season opens. The lake is declared to be "fished out" and is closed for one or more years until tests show it to be yielding good fish again. Heavy plantings of "eyed" eggs or fry are made to help the restoration.

These trial and error tactics have worked fairly well, and, over a period of twenty-five years, administrators have arrived at a fair estimate of the sustained annual yield of each lake. However, this type of administration does not provide the answers to several important questions. First, it is not possible to predict when a particular fishery may collapse, since it cannot be shown that any particular collapse is due to overfishing. Second, the optimum catches—the reaching of a proper balance between recruitment and mortality—are unknown and probably differ from the actual catches. Third, the effect, if any, of the plantings of "eyed" eggs or fry is not known.

It was partly from a desire to answer these questions that the investigation reported in this paper was begun; but, more important, these small lakes seemed to offer splendid opportunities to observe the effects of fishing pressure on populations which were known to be homogeneous. In these small lakes, the picture is not complicated by the possibility of immigration or emigration; the whitefish being caught in any lake are from a population that arose in that lake, is the sole whitefish population of the lake, and cannot be augmented by arrivals from outside populations or depleted by other than fishing or natural mortality. Thus many of the variables which complicate the study of a marine fishery, or fisheries of large lakes, are conveniently absent.

For the past seven years, the author, in co-operation with the provincial fisheries administration, has been taking samples of the commercially caught whitefish from a series of lakes and analyzing them for age composition and rate of growth. We hoped that we should be able to recognize overfishing or underfishing and assess the value of the whitefish hatchery. In this paper, I wish to discuss the whitefish populations of two lakes; Lake Wabamun and Pigeon Lake. Lake Wabamun has been closed for part of the period of study, whereas Pigeon Lake has been deliberately over-exploited. The whitefish populations of the two lakes thus form an interesting contrast, and clearly show the effects of fishing pressure. The lakes are similar in size (35-40 square miles) and have a maximum depth of 35 feet. The common sucker (*Catostomus commersonnii*) is the only whitefish competitor in both lakes. Both lakes are fished with gill nets of 5½ inch (stretched measure) mesh. Samples of the commercial catch have been taken twice each year; approximately three thousand fish have been measured and their ages determined. Calculations of the lengths of the fish at the end of each year of life were made by Van Oosten's method (1923).

THE LAKE WABAMUN FISHERY

The catches of whitefish from Lake Wabamun were from 150,000 to 200,000 pounds annually from 1918 to 1935 (Table 1); the years 1935 to 1940 saw greatly increased yields—up to 600,000 pounds, some three times as great as the average for the previous eighteen years. In 1941 the fishery collapsed; in spite of high effort, a low yield of small fish resulted. (see Table 1). As a consequence, the lake was closed for two years. In 1944 it was reopened and the catch has risen to a high

TABLE 1
TOTAL CATCH, TOTAL NETS, NUMBER OF MEN ENGAGED AND CATCH PER NET-MAN
IN LAKE WABAMUN

Year	Catch (lbs.)	No. Nets	No. men	Catch per net-man
1918	146,000	180	46	17.3
1919	134,500	96	35	40.0
1920	no data	—	—	—
1921	188,500	110	25	68.4
1922	141,000	202	46	15.2
1923	187,400	306	54	11.3
1924	309,500	202	63	24.3
1925	273,200	341	98	8.2
1926	172,900	411	107	3.9
1927	55,400	366	85	1.8
1928	74,300	365	66	3.1
1929	213,700	282	72	10.5
1930	155,700	384	64	6.3
1931	206,000	408	67	7.5
1932	189,300	486	81	1.8
1933	194,900	510	85	4.5
1934	no data	—	—	—
1935	241,000	846	207	1.4
1936	387,000	808	118	4.1
1937	653,200	983	983	0.7
1938	599,700	1439	1439	0.3
1939	521,600	3681	862	0.2
1940	334,200	1186	643	0.4
1941	174,300	764	382	0.6
1942	closed	—	—	—
1943	closed	—	—	—
1944	106,000	422	422	0.6
1945	318,900	791	791	0.5
1946	349,820	997	997	0.4
1947	532,780	1299	1299	0.3

level again. In the absence of age and growth data, it is impossible to state whether the collapse in 1941 was or was not due to overfishing.

Samples of the fishery were first taken in 1942 and have been taken regularly since. The age composition of these samples is shown in Table 2. In the year following the collapse of the fishery, the sample contained 69 percent of four-year-old fish; during the next four years, the samples contained more and more older fish—by the fall of 1947, only one percent of the sample was of four-year or younger fish.

TABLE 2
AGE COMPOSITION OF SAMPLES OF WHITEFISH FROM LAKE WABAMUN

Date	Size of sample	Percent of each age							
		2	3	4	5	6	7	8	9
January '42	34	0	0	69	23	8	0	0	0
Aug. '42-Feb. '43	73	0	5	43	43	7	2	0	0
June '43-Mar. '44	633	0	3	48	39	10	0	0	0
July '44-Jan. '45	251	0	0	14	54	28	4	0	0
July-Aug. '45	201	0	0	0 5	16	58	23	2	0 5
October '46	100	0	1	0	4	66	23	4	2
February '47	100	0	7	22	32	25	9	5	0
Sept.-Oct. '47	371	0	0	1	3	19	49	22	5

1763

In Table 3 the rates of growth of the year classes in the samples are shown; year classes from 1935 to 1943 are represented. During the period of closure and light fishing, there has been no change in growth rate. The apparent increase in growth of the later year classes (1942-1943) is mainly due to selective action of the gill nets.

Lake Wabamun shows us, then, that a fishery which is recovering from heavy exploitation has, first, a steady growth rate, and second, an increasing number of older fish in the population.

TABLE 3
CALCULATED TOTAL LENGTH (MM) OF EACH YEAR CLASS AT THE END OF EACH YEAR OF LIFE IN LAKE WABAMUN

Year Class	No. Fish	1	2	3	4	5	6	7	8
1935	2	100	203	305	364	400	421	438	—
1936	4	104	205	292	346	383	406	431	—
1937	33	106	209	288	344	380	410	432	449
1938	134	109	211	291	348	378	409	429	—
1939	134	112	210	294	347	383	406	428	461
1940	102	109	207	288	345	376	400	423	—
1941	21	112	202	269	333	370	405	—	—
1942	4	123	227	321	364	379	—	—	—
1943	5	112	222	311	—	—	—	—	—

THE PIGEON LAKE FISHERY

In the absence of growth and age composition data prior to the collapse of the Lake Wabamun fishery in 1941, it is impossible to say whether or not it was due to overfishing. It was decided to allow increased fishing in Pigeon Lake and follow the changes in the population which preceded a collapse of the fishery. Pigeon Lake has had a fairly steady yield of 150,000 to 200,000 pounds per year since 1918 (Table 4). In 1941 the catch was increased to nearly 600,000 pounds,

TABLE 4
TOTAL CATCH, TOTAL NETS, NUMBER OF MEN ENGAGED AND CATCH PER NET-MAN
IN PIGEON LAKE

Year	Catch (lbs.)	No. nets	No. men	Catch per net-man
1918	78,700	195	49	8.2
1919	144,000	64	60	37.5
1920	no data	—	—	—
1921	138,000	330	60	7.0
1922	152,000	290	75	7.0
1923	183,720	316	87	6.7
1924	228,700	700	102	3.2
1925	277,600	700	129	3.1
1926	248,200	995	162	1.5
1927	235,700	858	142	1.9
1928	144,800	525	167	1.6
1929	146,000	570	95	2.7
1930	148,000	684	114	1.9
1931	130,900	810	135	1.2
1932	196,300	810	135	1.8
1933	214,900	1128	188	1.0
1934	no data	—	—	—
1935	182,800	374	374	1.3
1936	135,300	284	284	1.7
1937	115,800	169	169	4.0
1938	179,000	451	262	1.5
1939	203,100	1405	265	0.5
1940	269,500	1405	228	0.8
1941	582,900	878	421	1.6
1942	354,600	633	293	1.9
1943	340,000	534	487	1.3
1944	485,000	354	260	5.3
1945	411,000	1095	826	0.5
1946	350,000	1232	1032	0.3
1947	160,000	798	798	0.3

over three times the previous annual average. This increased yield continued for six years; the fishery then collapsed in 1947 and the lake is now closed. It is interesting, and perhaps significant, that Lake Wabamun also collapsed after the same period, six years, of heavy fishing.

In Table 5 the age compositions of samples of the catch since 1942 are shown. In 1942 the average age was 5.1 years and ninety percent of the sample was of four-year-olds or older. With the increased fishing, the samples contained more and more younger fish; by the fall of 1947 nearly three-quarters of the catch was of two-year-olds and the fishery had collapsed.

TABLE 5
AGE COMPOSITION OF SAMPLES OF WHITEFISH FROM PIGEON LAKE

Date	No. fish	Percent of each age								
		1	2	3	4	5	6	7	8	9
Aug. '42	100	0	8	2	17	28	37	8	0	0
Dec. '43-										
Feb. '44	330	0	0	27.3	25.5	30.3	13.2	3.7	0	0
Aug. '44	127	.8	0	1.6	33	26	26.8	11	.8	0
Sept. '45	102	1	0	11.8	52.9	23.5	8.8	2	0	0
Dec. '46-										
Jan. '47	301	0	13.3	21.6	40.2	13.6	7.6	3.7	0	0
Oct. '47	215	0	74.5	20.9	3.3	0.4	0.4	0.4	0	0

1175

In Table 6 the rates of growth of each year class from 1935 to 1945 are shown. It is clear that with the increased fishing pressure, the growth rate has also increased. For example, the calculated length of a three-year-old increased from 285 mm. for the year class of 1935 to 333 mm. for the 1942 year class. The picture is more easily seen by examining Table 7. This table shows the weight of 100 whitefish for each year of the fishery since 1942; the weights were calculated by multiplying the average weight of each age group by the percent abundance of that age group in the sample. Table 7 also shows the average increments in total length for the whole population beyond the 1942 level, and the percent of the catch made up of fish four years old and younger. Maximum efficiency—greatest weight per hundred fish—was reached in 1944, the third year of the increased fishing, while maximum growth was not reached until two years later. The fast

TABLE 6
CALCULATED TOTAL LENGTH (MM.) OF EACH YEAR CLASS AT THE END OF EACH
YEAR OF LIFE IN PIGEON LAKE

Year Class	No. Fish	1	2	3	4	5	6	7	8
1935	9	102	203	285	332	371	396	415	—
1936	21	101	196	280	330	367	391	432	—
1937	23	106	215	306	349	381	421	447	—
1938	29	103	208	295	353	391	419	444	473
1939	69	108	210	308	364	403	444	459	—
1940	83	111	213	310	372	417	431	439	—
1941	32	114	222	318	383	403	—	—	—
1942	27	120	229	333	375	389	—	—	—
1943	32	118	232	330	363	—	—	—	—
1944	39	119	255	314	—	—	—	—	—
1945	24	122	252	—	—	—	—	—	—

TABLE 7
AGE COMPOSITION OF CATCH, WEIGHT (OUNCES) OF 100 FISH FROM THE CATCH
AND AVERAGE ANNUAL GROWTH INCREMENT (MM. TOTAL LENGTH), PIGEON
LAKE, 1942-1947

Year	Percent of catch 4 years and under	Calculated weight of 100 fish	Average annual growth increment
1942	27	3014	0
1943	52 8	2881	13 2
1944	35 4	3850 3	22
1945	65 8	3555	41
1946	75 1	2860 4	49 2
1947	98 7	1035 2	17 8

growth of 1946 failed to compensate for the youth of the catch, which was of 75 percent four-year-fish and younger.

We may conclude from the Pigeon Lake study that:

- (1) The fishery collapsed due to overfishing.
- (2) The overfishing resulted in increased growth rate.
- (3) At first, the increased growth rate more than compensated for younger fish in the catch, i.e. the average weight of the fish increased. After three years, the increasing number of young fish "caught up" to the increasing growth rate, i.e. although the growth rate increased for

another two years, the efficiency began to decrease as the fish were taken too young. This decrease in efficiency took place when more than 35 percent of the catch consisted of fish four years old and younger.

(4) In the final year—the year of collapse—the rate of growth decreased.

CATCH PER UNIT EFFORT

In these small whitefish lakes, the total catch is determined arbitrarily; also there is no control over the number of licenses sold. As a consequence, catch per unit of effort fluctuates inversely with number of licenses and the price of whitefish.

Tables 1 and 4 show the total catches, total men fishing, total nets and catch per net-man in Lakes Wabamun and Pigeon since 1918. In general, it may be seen that catch per net-man is low when the catch is high and vice versa. During the last three years in Lake Wabamun, catch per net-man has been very low; yet from the age composition data we know that the fishery is in good order and that there is no immediate danger of collapse. In Pigeon Lake, the lowest catches per net-man (0.5 and 0.8 in 1939 and 1940) preceded the largest catch in the lake's history, nearly six hundred thousand pounds in 1941. Much more precise statistics than numbers of men and numbers of nets are needed to give catch-effort data of any value. Even if better statistics were available, it is doubtful if they would help much in these fisheries; it is a common observation among fishermen that these lakes yield heavily per net right up to, and including, the season before a collapse. It would appear that catch per net is more a measure of availability than of abundance.

THE EFFECT OF THE HATCHERY

The question arises, may the collapse of a fishery be avoided by the planting of "eyed" eggs or fry? The study of Pigeon Lake makes this appear unlikely. For it has been noted that in 1942, most of the fish spawned in Pigeon Lake as four-year-olds—at the end of their fifth summer. In 1946, all the two-year-olds were spawners; with the increased growth rate, younger and younger fish became mature. The collapse of the fishery was not due, therefore, to the failure of the fish to reproduce. There are plenty of one- and two-year-old fish in the lake now. These observations are perhaps enough to answer the question but we have tried to prove the answer by comparing the strengths of year classes which have had hatchery support with those which have

not. A more complete account of this work has been published elsewhere (Miller, 1946).

Table 8 shows a comparison of the year classes of 1940 and 1941 in four lakes. The figures in the table show the percentage of fish of each year class found in the samples during six years of sampling. Note that the year class of 1940 is stronger in each lake whether or not it received support from the hatchery. In Tables 9 and 10, the strengths of various other year classes are compared; again note that year class strength appears independent of hatchery operations.

There is no evidence, then, that the hatchery can influence production.

TABLE 8
THE RELATIVE STRENGTHS OF THE WHITEFISH YEAR CLASSES OF 1940 AND 1941
IN FOUR ALBERTA LAKES

Lake	Year Class	Hatchery plant	Percent of catch at each age						Total
			2	3	4	5	6	7	
Pigeon	1940	5 million	8	27.3	33	23.5	7.6	0.4	99.8
	1941	none	0	1.57	52.9	13.6	0.4	—	68.5
Buck	1940	none	7.7	45	33.7	40	—	—	126.4
	1941	1 million	0	3.5	10	—	8.4	—	21.9
Lesser Slave	1940	41 million	0	11.9	41.7	37.2	—	—	90.8
	1941	30 million	0	10.5	20.8	—	4.0	—	35.3
Wabamun	1940	5 million	0	3.35	11.7	15.4	45.5	—	76.0
	1941	6 million	0	0	0.5	18	18.6	—	37.1

TABLE 9
THE RELATIVE STRENGTHS OF FOUR WHITEFISH YEAR CLASSES IN PIGEON LAKE

Year Class	Hatchery plant	Percent of catch at each age			Total
		4	5	6	
1938	4 million	17	30.3	26.8	74.1
1939	3 million	25.5	26.0	8.8	60.3
1940	5 million	33.0	23.5	7.6	64.1
1941	none	52.9	13.6	0.4	66.9

TABLE 10
THE RELATIVE STRENGTHS OF TWO YEAR CLASSES OF WHITEFISH IN LESSER
SLAVE LAKE

Year Class	Hatchery plant	Percent of catch at each age						Total
		5	6	7	8	9	10	
1936	none	27.2	50	5.2	7.8	0.4	0	90.6
1937	88.7 million	33.4	17.9	7.0	4.4	—	0	62.7

DISCUSSION

The data derived from the study of the fisheries of Lake Wabamun and Pigeon Lake make a number of conclusions possible, most of which are already known from research on marine fisheries. (For information on marine fisheries, I have leaned heavily on E. S. Russell's excellent little book, *The Overfishing Problem* (1942)).

- (1) In a fishery in which heavy exploitation has ceased, the average age of the population increases and the growth rate is slow and steady; (Lake Wabamun).
- (2) In a heavily exploited fishery, the average age of the population decreases and the growth rate increases; (facts clearly demonstrated over twenty-five years ago by Petersen (1922) for the plaice fisheries of the Kattigat, Belts and Baltic). We have seen that this increased growth rate at first more than compensates for the decreasing average age, i.e. the average weight of fish in the catch increases; but as the number of young fish in the catch increases, the increased growth rate fails to compensate and the average weight of fish in the catch begins to fall. This may occur before catches begin to fall off, so we may conclude:
- (3) Overfishing is not just decreased yield in the face of increasing effort; *it begins when decreasing average age overtakes increasing growth rate and before decreased yields are evident.* Another important conclusion is a corollary to this one:
- (4) The fishery will not maintain production at maximum growth rate; in order to produce sufficient thinning of the population to permit maximum growth rate, fish of less than commercial size had to be taken, and the fishery collapsed. Curiously enough:
- (5) At the time of collapse, the growth rate decreased. This may

be due to one- and two-year-olds surviving in enormous numbers when the older age groups were removed, thus producing a crowding effect.

- (6) The collapse of the fishery was not due to brood failures and hatchery plants can do nothing to prevent such collapses or assist in recovery following a collapse. (In an earlier paper on these same fisheries, (Miller, 1947), I was much impressed by the ability of the whitefish to mature at younger and younger ages as the older fish were removed. It was evident that fishing with selective gear for large fish could not deplete the brood stock. But I was incorrect in suggesting that, in such a fishery, overfishing was not likely to occur).

Of recent years, there have been attempts to devise a formula or formulae by the use of which one may determine when to harvest a year class in order to get its maximum bulk; i.e. the optimum catch. Such a formula must contain expressions for natural mortality rate and fishing mortality rate, which diminish the population, and also for rate of growth and recruitment which add to the population bulk. Ricker has done much interesting and valuable work along these lines which he has brought together in a recent important paper (Ricker, 1948). The effects of overfishing in Pigeon Lake, which the present paper describes, serve to emphasize how difficult it is to evolve a satisfactory formula. In such a formula it is usually necessary to assume that the rates of growth, of recruitment and of natural mortality are reasonably stable. For example, Ricker (1945) has calculated the best minimum size for bluegills in Muskellunge Lake, Indiana; he has shown how this size will change with different degrees of fishing (p) but has retained the same value for natural mortality rate in most of his calculations. Using Jackson's method (1939) of determining the average ratio of the number of individuals of each age to the number one year younger, I have calculated the annual mortality in Pigeon Lake to vary over a range of fifty percent from 1942 to 1946. Gill net returns are admittedly unreliable for such a calculation, but the large variation does suggest that the rate of natural mortality is not stable but varies perhaps with fishing pressure.

Rate of recruitment, too, may well be a function of fishing pressure. Fish of recent age groups in Pigeon Lake appear to be so numerous that their growth rates are decreasing. This abundance was probably

caused by a greater survival beyond the fry stage due to the removal, by fishing, of the older age groups.

That rate of growth is not constant, but varies with population density, is well known and formulac for the optimum catch try to take account of this fact. This is not easy to do, however. In Pigeon Lake, I have calculated that the increasing rate of growth during the six years of overfishing caused each age group to be 19.3 percent heavier each year. This same rate of fishing caused the age composition to change in such a way that 35 percent per year was added to the group four years old and less in age. The yield to the fishermen in weight of fish was the resultant of these two rates; the increasing rate of growth increased the yield for two years and then the increasing youth of the catch (which itself caused the increasing growth rate) began to decrease the yield. The resultant of the two rates, therefore, was a collapse of the fishery after six years of overfishing.

It would appear, then, that all the factors which must be included in a calculation of the optimum catch are quite likely subject to variation with varied fishing pressure. Furthermore, in the lakes under discussion, the whitefish has no serious competitor, which, in the event of the depletion of the whitefish, might occupy its place in the economy of the lake. This would not be the case in most fisheries; the threat of the ascendancy of undesirable species is a real one in many fisheries, (e.g. The Great Lakes). And so we have still another unpredictable effect of fishing pressure which will complicate calculations of the optimum catch.

A crude calculation of the extent of overfishing and an estimate of the optimum catch in Pigeon Lake may be made as follows:

The average annual catch during the six years of overfishing was 421,000 pounds. Since it took six years for this rate of exploitation to produce a collapse, we may conclude that each year one-sixth too much was taken. Therefore, the proper (optimum?) annual catch should have been six-sevenths of 421,000 pounds or 361,000 pounds. If this calculation is correct, the annual average for the period 1918-1940 of 179,000 pounds was only half the possible sustained yield.

Of course, it is a rather drastic procedure to over-exploit a fishery until it collapses in order to discover what the annual catch should have been. It has taught us, however, that the age composition of the whitefish catch must not be forced below 35 percent four-year-old and younger fish, and that age compositions of this order give the greatest average weight of fish.

ACKNOWLEDGMENTS

I am happy to acknowledge the great assistance I have received from the Fisheries Branch of the Province of Alberta, particularly from Mr. H. B. Watkins, Superintendent of Fisheries; the project would have been impossible without his cooperation or the help of his staff of field officers who did most of the measuring and scale collecting.

REFERENCES

- Jackson, C. H. N. 1939. The analysis of an animal population. *Journal of Animal Ecology* 8: 238-246.
- Miller, R. B. 1946. Effectiveness of a whitefish hatchery. *Journal of Wildlife Management* 10(4): 316-322.
- Miller, R. B. 1947. The effects of different intensities of fishing on the whitefish populations of two Alberta lakes. *Journal of Wildlife Management* 11(4): 289-301.
- Petersen, C. G. J. 1922. On the stock of plaice and the plaice fisheries in different waters. *Rep. Danish Biol. Station XXIX*.
- Ricker, W. E. 1945. A method of estimating minimum size limits for obtaining maximum yields. *Copeia*, June 30, 1945 (2): 84-94.
- Ricker, W. E. 1948. Methods of estimating vital statistics of fish populations. *Indiana University Publications. Science Series* 15, 1948: 101 pp.
- Russell, E. S. 1942. The overfishing problem. *Cambridge University Press*. 130 pp.
- Van Oosten, J. 1923. The whitefishes (*Coregonus chuapeaformis*). A study of the scales of whitefish of known ages. *Zoologica, Scientific Contributions N. Y. Zoological Society* 11(17): 380-412.

STATISTICS OF A LAKE TROUT FISHERY

F. E. J. FRY

*Ontario Fisheries Research Laboratory
Department of Zoology, University of Toronto
and*

Research Division, Ontario Department of Lands and Forests

SINCE 1936 THE lake trout fishery of lake Opeongo, Algonquin Park, Ontario, has been followed by means of a creel census. This census was undertaken as an experiment in methodology as well as to give information on the lake trout *Cristivomer namaycush* (Walbaum). Thus one of the main objects has been to determine to what extent a record of the fish caught, and of the circumstances surrounding their capture can yield a picture of the nature of the population under exploitation and of the changes that may take place in it. From time to time sections of the data collected have been analyzed, chiefly those relating to size and rate of capture (Fry and Kennedy 1937, Fry 1939^b, Fry and Chapman 1948). In this report the growth rate, age composition and fecundity are related to the statistics of catch with the express aim of building up from these data some logical picture of the lake trout population from which this fishery is drawn. Thus the object is to determine within whatever limits it may be possible to attain, the size of the population, its annual increment in weight, its spawning strength and the degree to which the fishery exploits it.

The procedure by which the Algonquin Park Creel Census is conducted and the forms used are described elsewhere (Fry 1939^b). Opeongo lake is the main census station in Algonquin Park and it has been possible to obtain an almost complete record of the fishing in this lake since virtually all the people visiting it, except for transient canoe parties, reach it by a single road and come in contact with the census taker.

Many of the fish taken are brought to the landing at Sproule bay either by guests of the fishing lodge located there or by parties who make a special trip for a day's fishing. Most of the fish brought to the landing are available for measurement and examination. The census worker guts these fish in return for the privilege of examining them and in this way material is accumulated for stomach analysis and fecundity studies as well as lengths, weights and scale samples.

Lake Opeongo is situated in the Precambrian shield at 45° 42' N. 78° 23' W. It has an area of approximately 20 square miles and a maximum depth of about 175 feet. In one respect Opeongo has not been an ideal lake for this study. It is a unit in name only since it consists of four basins isolated from each other by shallow and constricted channels so that each basin has its own limnological peculiarities and perhaps its own discrete population of lake trout. In any event it is certain that lake trout are present in all the basins in summer and it is highly unlikely that they mix during that season. In consequence, therefore, all sections of the population are not affected equally in all years by the fishing that goes on. Fashions change, good luck in one of the arms at one time will swing the majority of the effort to that basin, perhaps for the rest of the season, and may leave the populations of the other basins relatively untouched for the time being. This was especially true during the war years when both gasoline and leisure were restricted and the anglers did little exploratory fishing. Because of the distinctness of these basins the Opeongo fishery, limnologically speaking, should be considered the yield of four small lakes rather than that of a single lake of moderate size. However, although these considerations may greatly affect specific conclusions regarding action to be taken in respect to the Opeongo fishery, they appear to have no great bearing on the general conclusions presented here.

ACKNOWLEDGMENTS

The Algonquin Park Creel Census was initiated in 1936 at the suggestion of Dr. W. J. K. Harkness at the time Director of the Ontario Fisheries Research Laboratory of the Department of Zoology, University of Toronto. The expense of the work was borne by this laboratory in the first year aided by the Department of Lands and Forests through the kind support of Mr. F. A. MacDougall. Subsequently a grant in aid of the census was received from the National Research Council of Canada through the National Committee of Fish Culture. In 1946 the work was transferred to the Research Division of the Department of Lands and

TABLE 1

CHARACTERISTICS OF THE LAKE TROUT CATCH REMOVED FROM LAKE OPEONGO IN THE YEARS 1936-1947 INCLUSIVE. THESE ESTIMATES ARE BASED ON THE ASSUMPTION THAT 80% OF THE TOTAL CATCH WAS RECORDED BY THE CREEL CENSUS.

Year T	Units effort (100 hrs.) $E(T)$	Estimated number of fish caught	Removal pounds	Pounds per acre	Average Age fish landed years	Availability number captured per 100 boat hours $C(T)$
1936	20.3	2600	9400	0.7	—	128
1937	22.4	2700	7450	0.56	7.98	121
1938	16.3	1650	3940	0.29	7.11	101
1939	13.8	1550	4350	0.32	7.65	112
1940	11.7	1400	3320	0.25	7.11	120
1941	11.3	1100	2520	0.19	6.94	97
1942	5.7	630	1550	0.12	7.12	111
1943	7.1	900	2330	0.17	7.74	126
1944	9.2	1050	3390	0.25	8.13	126
1945	14.0	1420	5600	0.42	9.03	101
1946	17.4	1220	3620	0.27	9.10	70
1947	12.3	885	2500	0.19	7.81	72

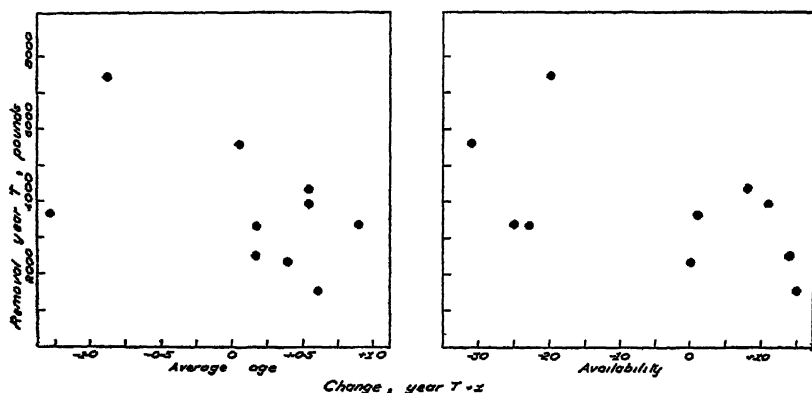
Forests of Ontario, which agency is now continuing it. The field work in the various years has been the responsibility of junior members of the staff of the Ontario Fisheries Research Laboratory and latterly of the Research Division of the Department of Lands and Forests. Among these were the following: Messrs. W. A. Kennedy, J. R. Brett, H. M. M. Tovell, D. W. Kirk, J. Long, R. R. Tasker, N. S. Baldwin and N. V. Martin. The writer is also indebted to Dr. D. B. DeLury and to Mr. D. Teichroew for much guidance and assistance in the mathematical analysis of these data. Mr. Teichroew performed most of the computing, as well as acting as mathematical consultant during the preparation of the paper. I am also indebted to Dr. J. M. Speirs for a critical reading of the manuscript.

STATISTICS OF REMOVAL

Introduction

The general statistics of removal of lake trout from lake Opeongo are given in table 1. During the period covered by the creel census the estimated number of lake trout captured per year has varied from 2700

FIGURE 1.



THE RELATION BETWEEN REMOVAL IN YEAR T AND THE CHANGE IN CHARACTERISTICS OF THE CATCH IN YEAR $T + 1$ THE COEFFICIENTS OF CORRELATION ARE: WITH CHANGE IN AGE $r = 0.554$; WITH CHANGE IN AVAILABILITY $r = 0.560$ THESE ARE ON THE BORDERLINE OF THE 0.05 LEVEL OF PROBABILITY. THE SCATTER IS PRESUMABLY DUE TO VARIATION IN THE STRENGTH OF THE INCOMING YEAR CLASSES.

in 1936 and 1937 to 630 in 1942. This variation has been the result of changes both in the size of the population and in the intensity of fishing effort. Previous to 1936 lake Opeongo was a relatively inaccessible place. It could be reached by car only by travelling a road which was the converted right of way of an old logging railway and this approach was from the more sparsely settled eastern section of the province. In 1936 a highway to the west was opened with the result that many more anglers from southern Ontario and the United States visited the lake, thus greatly increasing the rate at which the Opeongo trout population was exploited.

In each of the years 1936 and 1937 approximately 2000 boat hours of fishing effort were expended in the pursuit of lake trout in Opeongo. Lake trout in this district are typically taken by a troll on a metal line; in general two lines are fished from a boat at the same time. Fishing intensity waned somewhat in 1938 and dropped off still further with the beginning of the war. In 1942 the fishing effort expended was only about 600 boat hours, the least recorded. After the war the expenditure of fishing effort returned to approximately the 1938-39 level.

The estimated poundage of lake trout removed has varied from 1550 to 9400 pounds per year. These values represent an annual removal of from 0.12 to 0.7 pounds per acre of water surface. Although the removals are of a low order of magnitude in terms of yield per acre, they have had an influence on the fishery as figure 1 indicates. This graph shows the

correlation between removal in one year and two indices of the fishery in the following year.

Provisionally at least, these two correlations may be taken to indicate that fishing mortality is not negligible in comparison with natural mortality.

Age Composition of the Catch

A detailed estimate of the removal by age groups is given in table 2.

TABLE 2
ESTIMATED NUMBERS, $K(x;T)$, OF LAKE TROUT REMOVED AT VARIOUS AGES FROM LAKE OPEONGO DURING THE YEARS 1936 TO 1947. SEE TABLE 1 FOR TOTALS

Year of Capture	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII
1936	30	95	128	233	474	665	478	260	118	57	19	25	10	15	11
1937	0	4	34	198	650	1025	555	176	38	4	8	8	0	0	0
1938	12	74	127	275	420	439	193	90	3	9	8	0	3	0	0
1939	39	36	116	221	321	393	223	90	47	24	13	15	4	4	4
1940	20	84	82	224	434	364	120	46	14	6	0	6	0	0	0
1941	8	79	144	275	235	200	104	22	11	11	11	0	0	0	0
1942	7	18	46	117	217	121	53	28	8	9	2	2	0	0	0
1943	6	42	42	121	272	211	133	42	0	24	6	0	0	0	0
1944	8	26	84	114	197	202	198	93	44	31	9	22	9	4	4
1945	0	11	32	69	170	352	373	159	84	37	21	43	16	37	16
1946	19	30	78	116	240	325	217	93	47	26	11	7	0	7	4
1947	3	30	55	85	217	221	153	76	18	12	3	0	6	3	3
Mean	11	40	76	165	306	350	212	83	20	18	8	9	3	5	3
Percentage	0.8	3.0	5.8	12.5	23.2	26.6	16.1	6.3	2.2	1.3	0.6	0.7	0.3	0.4	0.2

These estimates were made by determining the age composition of a sample of scales gathered as the opportunity presented itself throughout the fishing season. These scale samples represented 30% of the estimated catches. It is presumed that there was no bias in the collection of these samples although it is certain that this assumption is not strictly true, for as can be well imagined larger than average fish were more likely to be brought to our attention than smaller ones. However, since so large a sample was examined the effect of this bias cannot be very great. In the case of the estimate for 1936 this procedure was not followed since some of the scale samples were lost. The age composition of that year's catch was estimated from the age composition of each size class as estimated from the average growth in length over the eleven year period.

The Opeongo lake trout first enter the fishery at age III but only a negligible number of this age group are taken. Age groups V to X have provided 90% of the total fishery, ages VII and VIII making the most important contribution within this range. These two age groups have been responsible for 50% of the total fishery. The highest age that has been read is XVII. Fish of ages XV to XVII were taken only infrequently after 1936 except in 1945 when a markedly high number of old fish were taken. These no doubt were from the less accessible basins of the lake which were not so highly exploited during the war years.

The Virtual Population

When the total catch and its age composition are known for a number of years it is possible to sum up the contribution of each year class which has passed through the fishery in that time. This summation of the year class has to remain in abeyance until the year class has passed completely through the fishery so that no more of its members will be captured to add to the total. It is proposed that this complete contribution of a year class to the fishery shall be termed the *virtual population*. It is something which can be seen, piecemeal of course, but which is not the true population. However it is all that we are able to see of most populations of fish and it places one limit on a number of estimates. Moreover the value of these estimates is the greater the more intensive the fishery and hence the greater the need for giving it close statistical attention.

Estimates of the virtual population for various year classes of lake trout at each age as they have passed through the Opeongo fishery since 1936 are given in table 3. The values for the virtual populations in table 3 were obtained by using formula 1 (page 66) which is equivalent to summing diagonally the values for removal by ages given in table 2. For example, fish of the 1923 year class were captured at ages XIII to XVI in the years 1936 to 1939. No fish of this year class of age XVII were taken. The total catch of age XIII and older is $19 + 8 + 3 + 4 = 34$ fish and the virtual population at the beginning of age XIII is stated to be 34.

Lake trout enter the Opeongo fishery at age III and persist to age XVII, a period of fifteen years, so that with only 12 years' records it has not yet been possible to complete the observations on a single year class from the beginning of its entry into the fishery. However age groups XIII to XVII have in general been of minor importance and it is felt that the addition of the mean values of these groups where necessary to the 1931 to 1936 year classes has not introduced serious error.

TABLE 3

THE VIRTUAL POPULATION, $V(x;T)$, OF LAKE TROUT OF VARIOUS YEAR CLASSES T AT DIFFERENT AGES. BOLD FACE FIGURES ARE MEANS FOR THEIR COLUMNS

Year Class (Hatching Year)	Age x																
	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII		
$T - x$																	
1919																11	
1920															15	0	
1921														10	0	0	
1922													29	4	4	4	
1923												34	15	7	4	0	
1924										69	12	4	4	0	0	0	
1925									140	22	18	15	0	0	0	0	
1926								326	66	28	19	6	0	0	0	0	
1927							683	207	31	28	4	4	4	4	4	4	
1928						1396	781	176	86	39	33	22	20	20	16		
1929					1861	1387	362	167	77	63	52	50	50	41	4		
1930				1665	1432	782	343	120	74	63	54	48	26	10	3		
1931			1371	1243	1045	625	232	112	90	82	58	49	86	6	3		
1932		1294	1199	1165	800	569	205	101	73	73	42	21	14	8			
1933	1223	1193	1189	1062	841	407	207	154	112	68	31	20	11				
1934	1129	1122	1059	939	715	480	359	226	133	49	23	20					
1935	1277	1265	1229	1147	872	655	444	246	87	40	28						
1936	1388	1349	1265	1121	1004	732	530	157	64	46							

With the aid of the extrapolation mentioned above, estimates of the total contribution of four year classes (1933-1936) to the Opeongo lake trout fishery can be made. These are surprisingly small, varying from 1100 to 1400 fish, an average of less than one fish to ten acres of water surface. The contributions of the five preceding year classes, judging from their contributions since 1936, were somewhat larger, although probably no more than twice as great in any instance since the level of exploitation of the fishery was considerably lower before 1936 than it has been subsequent to that year.

Maximum Estimates of the Rate of Exploitation

The figures for the virtual population of lake trout at various ages given in table 3 represent minimum values for the number of fish at each age present in the lake at the beginning of a given fishing season, since all these fish which were of the age in question at that time were subsequently captured. Others of the same age were no doubt also present which subsequently died from causes other than the angler's troll.

An upper limit can therefore be set to the level of exploitation $K(x;T)/V(x;T) \times 100$ by calculating the percentage of the virtual

population removed in each fishing year. Let us take as an example the members of the 1930 year class removed in 1937, in which year they were age VII. In table 2 the estimate of the catch, $K(x;T)$, from this age group in 1937 is 650. In table 3 the estimate of the virtual population, $V(x;T)$, of this year class as of the beginning of year VII is 1432. The maximum value for the percentage of that year class which was removed in 1937 is therefore 45.4%.

Percentage removals were calculated in this manner for each value given for the virtual population in table 3 up to age XIII. These values were then averaged by age groups thus giving mean figures for the maximum level of exploitation. These mean values are plotted in figure 2 where they are referred to as the *virtual percentages captured*. It will be seen that these values are very low for ages III and IV, rise to a peak in the neighbourhood of ages VIII and IX and fall away again at the higher ages.

The Maximum Force of Fishing Mortality

The mean values for the maximum level of exploitation shown in figure 2 do not take the fishing effort into account. However if the premise be granted that for a given age group the yield per unit effort is proportional to the size of the population, then a maximum value can be derived for the force of fishing mortality by taking the virtual population to be the equivalent of the actual population. This assumes the natural mortality to be zero. In which case formula 11 (page 66) describes the decline of the population, and the maximum force of fishing mortality is obtained by inserting the appropriate numerical values in formula 11 (page 66). This force, $k(x)\text{max.}$, is expressed here as the fraction of the population removed per 100 boat hours of fishing effort.

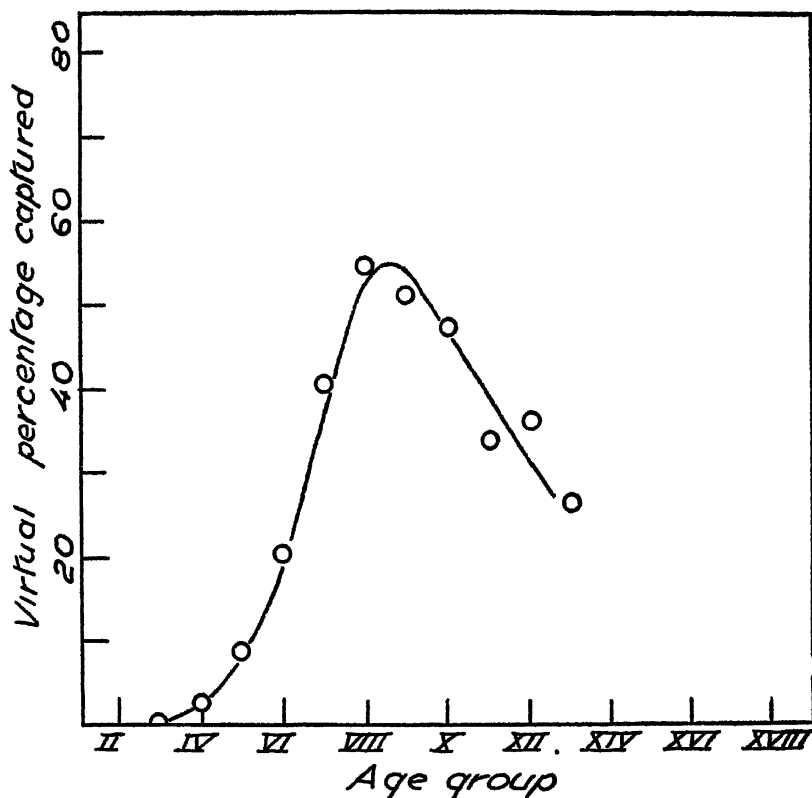
If we refer again to the example given above, the virtual percentage of 45.4% (0.454 when expressed as a fraction) was removed from the 1930 year class in 1937 by the expenditure of 22.4 units of effort. With the numerical values in the case cited, formula 11 becomes

$$1 - 0.454 = \exp [-22.4k(\text{VII})\text{max.}]$$

from which a value of 0.026 is obtained for $k(\text{VII})\text{max.}$ Values for $k\text{ max}$ calculated as above are given in table 4.

These values for the maximum estimate of the force of fishing mortality represent a mean response throughout the fishing season. The response of the population to the fishery is not all uniform within the period of the year when active fishing is being pursued, as is discussed

FIGURE 2.



THE PERCENTAGE OF THE VIRTUAL POPULATION CAPTURED AT DIFFERENT AGES OVER THE PERIOD 1937-1946. FOR A DEFINITION OF THE VIRTUAL POPULATION SEE PAGE 32. THESE VALUES REPRESENT MAXIMUM ESTIMATES FOR THE LEVEL OF EXPLOITATION OF THE OPEONGO LAKE TROUT POPULATION.

later (page 66). However, differences in fishing effort in different years are spread more or less uniformly over the whole season so changes in this mean value indicate changes in the condition of the fishery as a whole.

Removal of Year Classes

If the reduction in the number of members of a year class were completely known each year, it is obvious that plotting the cumulative loss in numbers against the cumulative percentage loss in terms of the original strength of entry would result in a straight line. The same is of

TABLE 4

MAXIMUM VALUES FOR THE FORCE OF FISHING MORTALITY, k_{max} , IN VARIOUS YEARS AT DIFFERENT AGES. THE VALUES REPRESENT MAXIMUM ESTIMATES OF THE MEAN FRACTION OF THE POPULATION REMOVED PER 100 BOAT HOURS OF FISHING EFFORT

Year Class	IV	V	VI	VII	VIII	IX	X	XI
1926								.038
1927							.084	.006
1928						.063	.043	.057
1929					.060	.047	.056	.068
1930				.026	.050	.076	.042	.098
1931			.008	.031	.072	.061	.019	.021
1932		.0013	.017	.033	.087	.064	.061	.000
1933	.0004	.0071	.017	.062	.059	.050	.045	.054
1934	.0036	.0083	.023	.035	.050	.062	.059	.072
1935	.0022	.0063	.024	.050	.054	.065	.072	.045
1936	.0053	.0119	.018	.067	.034	.086	.051	.026
Mean	.0029	.0070	.018	.0435	.0582	.0638	.0532	.0486

course true if the catch and the virtual population are known, in which case ignorance concerning natural mortality is exactly balanced by ignorance concerning the precise value of the force of fishing mortality. A plot of the virtual population for the 1932 year class is shown as the unbroken line in figure 3A. If such straight lines could be plotted before the actual size of the virtual population had been known, or in other words before the year class had passed completely through the fishery, they would have a prediction value since they could be extrapolated to give an estimate of the probable total catch from that year class.

Curves of this type can be constructed by the use of the mean values of k_{max} and the effort expended in the various years by again using formula 11 to determine the probable percentage of the virtual population taken in the year in question. This estimate will be fictitious in that it will not take into account the actual degree of catchability of the fish in that particular year, since a mean value will be used for the force of fishing mortality. However, it is assumed that over a few consecutive years fluctuations in catchability will be in both directions from the mean and that if such an estimate is low in one year, the estimate for the same year class another year may be equally high.

The broken line in figure 3A is a replot of the 1932 year class using the mean values for k_{max} . The degree of deviation from linearity is

Figure 1 consists of two graphs, A and B, plotting the cumulative number of fish captured (Y-axis) against the cumulative percentage of the virtual population entering age II (X-axis).

Graph A shows two lines representing the period 1935-1940. The solid line represents the 1935-1940 period, and the dashed line represents the 1940-1945 period. Both lines show a positive correlation, with the solid line generally higher than the dashed line.

Graph B shows multiple lines representing different years from 1940 to 1955. The lines are labeled with years: 1940, 1943, 1945, 1947, 1949, 1951, 1953, 1955. The lines show a positive correlation, with the 1955 line being the highest and the 1940 line being the lowest.

considerable but an estimate based on the four lower points would only give an underestimate of 20% for the final catch.

Other curves in which the mean value of k max were used to calculate the probable exploitation of the virtual population are presented in figure 3B. Arithmetically the derivation of the estimated percentages is a little round-about but not complicated. As an example let us take the first two terms for the 1932 year class. Fish of that year class were captured at age IV in 1936 and age V in 1937. From table 1 $E(1936) = 20.3$ and $E(1937) = 22.4$. From table 4 the means for the maximum force of fishing mortality are $k \text{ max (IV)} = 0.0029$ and $k \text{ max (V)} = 0.0070$. Therefore, the values for $\exp(-k \text{ max } E)$ for these two age groups from the 1932 year class are respectively 0.059 and 0.157, which yield values for $1 - K(x;T)/V(x;T)$ of .94 and .85. The value of .94 for $1 - K(x;T)/V(x;T)$ at age IV gives an estimate of 6% for the percentage removed from the virtual population attaining that age. Similarly the estimate of removal from the virtual population attaining age V is 15%; and since only 94% of those attaining age IV were estimated to attain

age V, then the further removal from the virtual population entering the fishery was $(15 \times 94)/100 = 14\%$. And the catch for ages IV and V represents a cumulative removal of 20% of the virtual population attaining age IV.

The points for a given year class shown in figure 3B scatter considerably and the trend is certainly not at all stabilized until about 50% of the potential catch has been removed and the year classes which have gone completely through the fishery did not all attain their final rank until the removal had been 80%. Thus as a means of prediction, these removal curves lack somewhat in precision. However, they distinguish between the best and worst year classes at about 40% removal. Moreover, it is also probable that the curve for the 1932 year class, which is the most out of order, sags at its lower end because in the years of the fishery when members of this year class were young, such small fish were often eaten on the beach instead of being recorded in the creel census.

Estimates of the probable rank of the year classes shown in figure 3, based on the catch to 80% removal, are given in table 11 where they are all compared with the 1938 year class. The yields from the earlier year classes measured were about one and two-thirds that of the 1937 year class. The estimate for the 1940 year class is about 80% of the yield of the year class hatched in 1938.

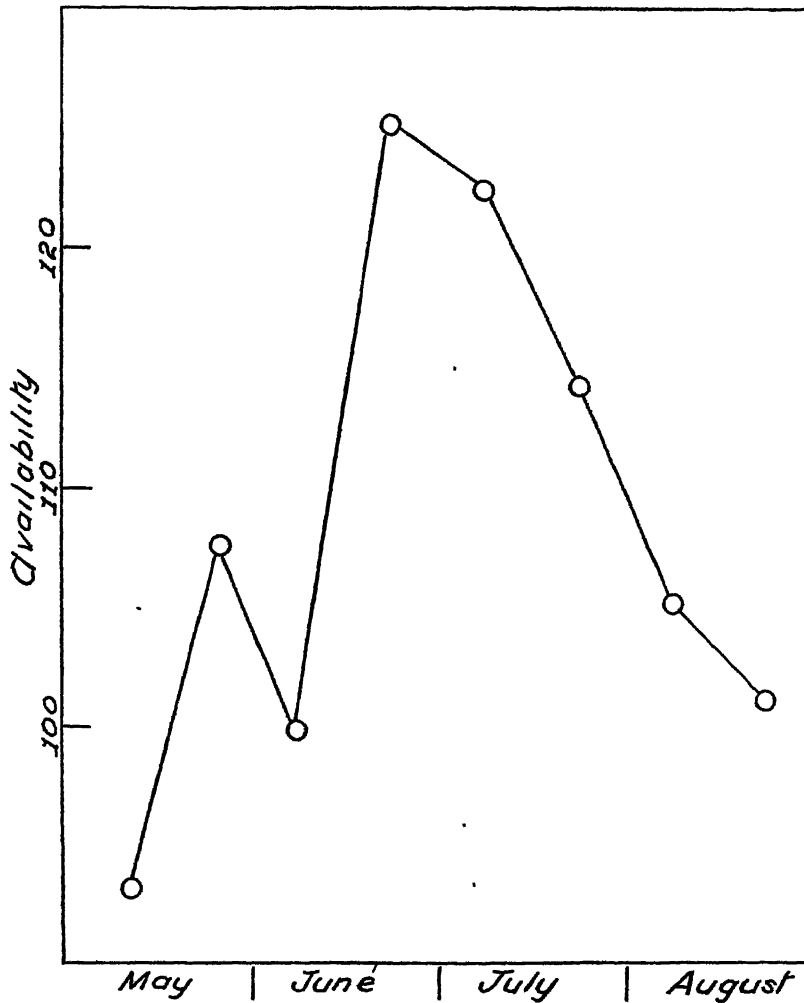
STATISTICS OF AVAILABILITY

Introduction

The yield per unit effort, $C(T)$, or availability of lake trout in lake Opeongo has varied from 128 fish per 100 boat hours to a low of 70. There is typically a seasonal cycle in the availability which is illustrated in figure 4. Typically lake trout are hard to catch in early May. They become somewhat more available in late May but there is a recession in fishing success in early June. Fishing is at its best in late June and early July. Subsequently fishing falls off progressively in late July and August. The trend in fishing in September is not shown in figure 4 since data were lacking for some of the years, but in general it improves somewhat over the August availability. Early October fishing is usually extremely poor. The season closes about October 15. This annual cycle in availability has been related to the migratory and feeding behaviour of the lake trout in response to the summer cycle of thermal stratification (Fry 1939^b).

Within a fishing season, therefore, there is no simple relation between the size of the population and the yield per unit effort. This is not sur-

FIGURE 4.



THE ANNUAL CYCLE OF AVAILABILITY OF LAKE TROUT IN LAKE OPEONGO, ONTARIO. THE CURVE IS BASED ON THE MEANS FOR THE YEARS 1936 TO 1947. AVAILABILITY IS EXPRESSED AS THE NUMBER OF LAKE TROUT CAPTURED PER 100 BOAT HOURS.

prising since the fishing takes place in those months when thermal conditions in the lake are changing continuously and are having a marked effect on the activity and distribution of the fish. #1 214

While the curve shown in figure 4 is the mean for the 10 year period it is not inevitable that the cycle take this form every year. Apparently

annual variations in hydrological conditions in lake Opeongo offer enough variety to make the annual response of the lake trout vary.

The Relation of Age to Availability

The availability of lake trout of different ages in lake Opeongo has been determined by proportioning the yield per 100 boat hours according to the age composition of the catch. These data are presented in table 5.

TABLE 5
MEAN ANNUAL VALUES OF AVAILABILITY, $C(T)$, OF LAKE TROUT OF DIFFERENT AGES IN LAKE OPEONGO IN THE YEARS 1937-1947, EXPRESSED AS THE NUMBER CAPTURED PER 100 BOAT HOURS. FOR ANNUAL TOTALS SEE TABLE 1.

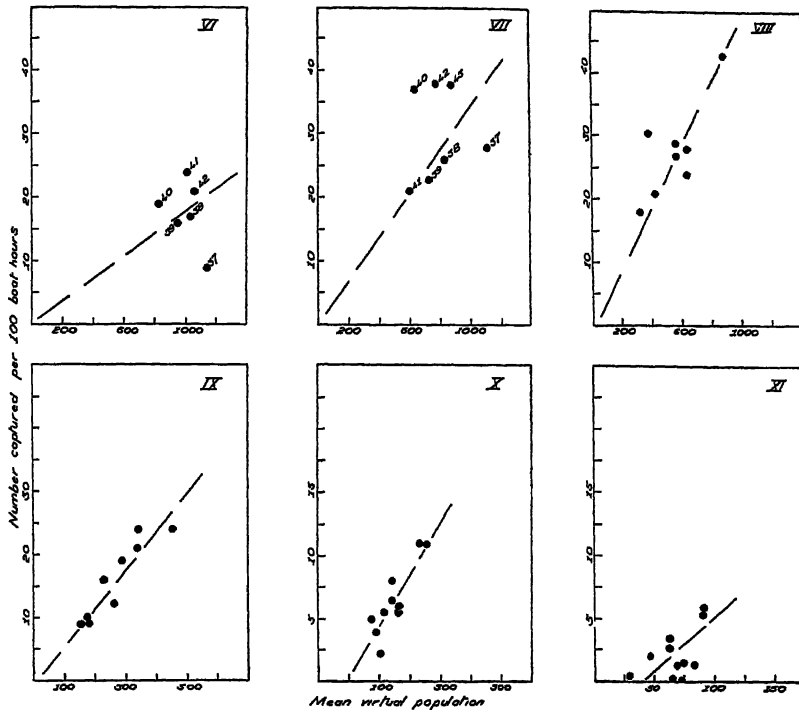
Year of Capture	Age																
	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII		
1937	0	0.2	1.5	8.9	28.1	45.9	24.8	7.9	1.7	0.2	0.3	0.3	0	0	0		
1938	0.8	4.5	7.2	16.8	25.6	26.9	11.9	5.5	0.2	0.6	0.2	0	0.2	0	0		
1939	2.8	2.6	8.2	15.7	22.9	28.0	15.9	6.4	3.3	1.7	0.9	1.1	0.3	0.3	0.3		
1940	1.7	7.2	7.0	19.2	37.2	31.2	10.3	3.9	1.2	0.5	0	0.5	0	0	0		
1941	0.8	7.0	12.7	24.1	20.6	17.7	9.2	2.0	0.1	0.1	0.1	0	0	0	0		
1942	0.2	3.1	8.2	21.0	38.2	21.4	9.3	5.1	1.2	1.6	0.1	0.1	0	0	0		
1943	0.8	5.9	5.9	17.6	37.8	29.4	18.5	5.9	0	3.4	0.8	0	0	0	0		
1944	1.1	3.2	10.1	13.8	23.4	24.4	23.9	11.1	5.3	3.7	1.1	2.7	1.1	0.5	0.5		
1945	0	0.7	2.2	4.7	15.3	24.1	25.5	10.9	5.8	2.6	1.5	2.9	1.1	2.6	1.1		
1946	1.1	1.7	4.5	6.7	13.8	18.7	12.4	5.4	2.8	1.5	0.6	0.4	0	0.4	0.2		
1947	0.2	2.5	4.4	6.9	17.5	17.8	12.4	6.2	1.5	1.0	0.2	0	0.5	0.2	0.2		

The availability of lake trout in lake Opeongo is highest at age VII, which is two years before the age at which k max is highest. The availability at age VIII is, however, almost equally high. Availability at the higher ages drops rapidly, presumably due to the drop in the size of the population.

The prime question with regard to availability is whether it bears any relation to the size of the population fished and what that relation may be. It has already been demonstrated (figure 4) that within a fishing season the relation of availability to the population must change. It remains to be seen whether the annual means have any simpler relation. The relation between the annual mean of availability at various ages to the virtual population at the same age is shown in figure 5.

As figure 5 indicates there are varying degrees of correlation between availability and the virtual population which range from none at all at age VI, since the negative trend displayed by these points can be considered absurd, to highly significant correlations at higher ages.

FIGURE 5.



THE RELATION BETWEEN AVAILABILITY AND THE SIZE OF THE VIRTUAL POPULATION AT VARIOUS AGES. NOTE THAT THE MEAN VIRTUAL POPULATION WITHIN THE SEASON IS USED, THE VIRTUAL POPULATIONS GIVEN IN TABLE 3 ARE ESTIMATES FOR THE BEGINNING OF THE FISHING SEASON. THE LINES TO INDICATE THE TRENDS WERE FITTED BY EYE.

Lack of correlation between availability and the size of the virtual population does not necessarily mean similar lack of correlation between availability and the true population, for the difficulty may be major fluctuations in natural mortality which would tend to make the values for the virtual population the more erratic of the two values correlated. However, the seasonal changes in availability shown in figure 4 cannot all be due to mortality. This conclusion is particularly strengthened by the fact that the lake trout in lake Opeongo undertake a summer migration which seems almost bound to influence their catchability as indeed all the facts known about their response demonstrate (Fry 1939^b). There is also some indication that in a given year the trend in the relation between availability and the size of the virtual population is in the same direction in both ages VI and VII. The years of capture are indi-

cated on the graph for the various points. The correlation, however, is not strong and more data are required before it can be considered that the trend is definite. It appears reasonable to conclude, therefore, that lack of correlation between availability and the virtual population is more likely to be due to fluctuations in the annual migratory cycle than to major fluctuations in natural mortality although this latter possibility cannot be entirely disregarded.

In those cases where there is a close correlation between availability and the virtual population no such problem of interpretation exists. It may be concluded with confidence that in these cases both the availability and the virtual population present stable reflections of the true population. Since the annual cycle of availability is exhibited by these older fish as well as the younger ones, it is to be presumed that the older fish present a more stable response to the annual migratory cycle and thus a more consistent annual cycle of availability.

A second interpretation that might suggest itself is that, since the size of fish appears to have a strong influence on their availability, annual differences in growth rate might have a marked influence on the availability of the entering age groups. This, however, does not appear to follow from our data as a comparison of tables 5 and 7 will demonstrate.

STATISTICS OF GROWTH

Introduction

The introduction of the method of making scale impressions by passing a strip of cellulose acetate with the scales outer side down on it through a jeweler's roller has greatly facilitated the preparation of lake trout scales for reading. We learned of this method through W. R. Martin of the Atlantic Biological Station. Our impressions were made on cellulose acetate 0.040 inches thick without heating. No cleaning of the scales was necessary. Before the scale impression method was adopted the scales were cleaned and mounted in glycerine waterglass.

Three to four scales were prepared and one read, the rest not being examined unless difficulty was encountered in the first one. All places where a complete circulus succeeded a series of incomplete ones were accepted as annuli. Reading of the scales was begun in 1938 but by far the greater number were read in 1947. I am indebted to Miss L. C. Craigie and F. P. Maher for most of the scale readings.

Growth in Length

Although the significant dimension from the point of view of produc-

tion is weight, the size age relationship was worked out primarily from measurements of body length. This course was adopted because of the greater ease with which measurements of length could be made in the field and especially because the fish were often dressed before the census worker had an opportunity of examining them. However, extensive series of weight determinations were also made so that it is believed that there is but little error in the age weight relationships deduced from the age length data.

To determine the degree of variation of the growth rate in the various basins, readings for fish of the VI, VII and VIII groups were separated into locality in cases where the place of capture had been noted. The means for these data are given in table 6. No substantial difference was found in the size attained at these three ages in any of the four basins. It is assumed that growth would also be similar at other ages.

TABLE 6
SIZE COMPOSITION OF CATCHES OF AGE GROUPS VI, VII AND VIII TAKEN IN THE YEARS 1937, 38, 39, 41, 42 43, 44, 45 AND 46 ARRANGED ACCORDING TO LOCALITY OF CAPTURE

Age	Locality	Fork length inches											Mean
		14	15	16	17	18	19	20	21	22	23	Length Inches	
VI	North Arm	3	9	26	10	2						16.0	
	South Arm	4	14	25	19	9	4					16.4	
	East Arm		19	37	38	15	5					16.6	
VII	North Arm			14	39	42	20	6	2	1		17.8	
	South Arm			18	38	51	29	3	1	0	1	17.9	
	East Arm			10	28	27	27	0	2	2		17.9	
	Annie Bay			5	15	19	12					17.7	
VIII	North Arm			2	6	43	40	29	11	1	1	19.0	
	South Arm			6	20	42	57	25	2			18.5	
	East Arm			3	11	43	46	24	2	2	1	18.8	
	Annie Bay				4	11	13	4	2			18.7	

There has been but little variation in the average size reached at a given age in the different years as table 7 indicates. What variation there has been does not appear to show any particular trend and would seem to reflect no more than annual variations in growing conditions

TABLE 7
THE EFFECT OF CALENDAR YEAR AND SEX ON THE AVERAGE LENGTHS ATTAINED
BY LAKE TROUT OF LAKE OPEONGO AT VARIOUS AGES

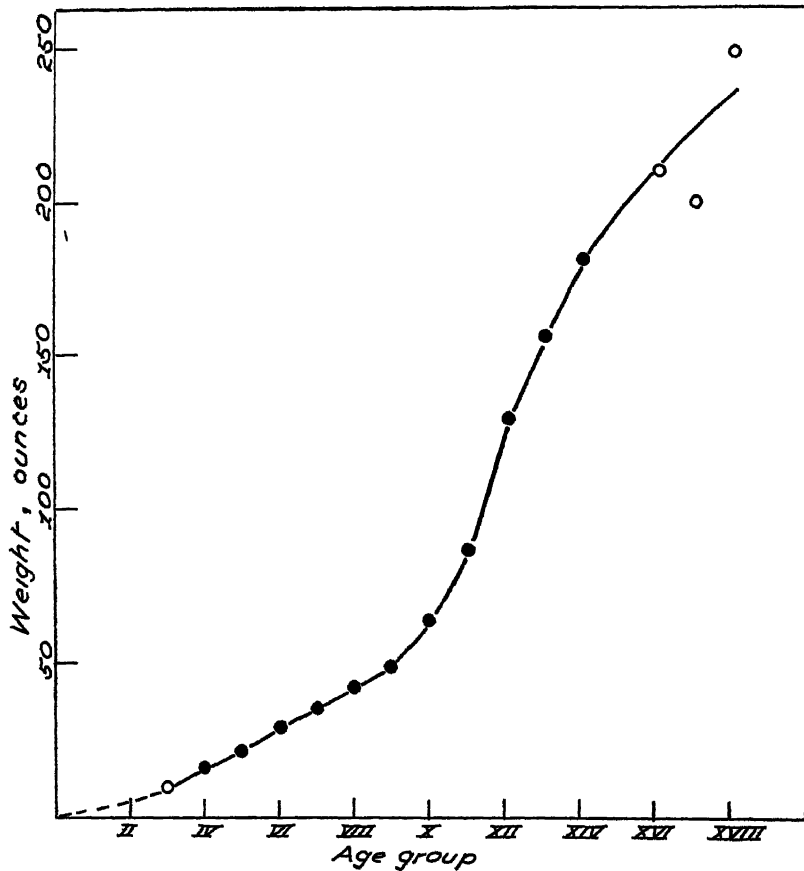
Year of Capture	Age					
	V	VI	VII	VIII	IX	X
1937	—	16 7	18 1	19 3	20 6	21 9
1938	14 1	16 7	17 9	19 2	20 1	21 8
1939	14.7	15 9	17 7	18 7	20 1	21 6
1940	14 6	16 3	17 9	19 2	20 2	22 3
1941	14 7	16 3	17 5	18 9	20 4	22 1
1942	14 7	16 8	18 3	19 1	20 2	21 2
1943	14 9	16 4	17 8	19 0	20 4	22 8
1944	14 7	16 7	18 2	18 8	19 8	20 6
1945	14 5	15 9	17 3	18 7	19 9	21 2
1946	14 5	16 0	17 5	18 6	19 7	21 6
1947	14 0	16 0	17 5	18 5	21 1	—
1943 } ♂	14 7	16 4	17 7	18 6	19 9	21 6
to						
1947 } ♀	14 7	16 2	17 6	18 8	20 3	21 2

There is no evidence of any progressive change in growth rate attendant on the increased exploitation of the fishery which has occurred since the opening of the Algonquin Park highway in 1936. It also appears that there is not appreciable difference in the growth rate of the two sexes. Table 7 shows a comparison of the average lengths attained by the two sexes at various ages. The relation between length and age found for all samples from 1937 to 1946 inclusive are shown in table 8.

The growth in length of the Opeongo lake trout as determined by our scale readings is considerably slower than had been inferred from the modes found in a sample gillnetted in 1936 (Fry and Kennedy 1937). The growth rate of the Opeongo fish is most similar to that reported by Juday and Schneberger (1930) for Wisconsin lakes being less than that found in the populations studied by Applegate 1947, Van Oosten 1943 and Royce M.S. The growth rate in Opeongo is, however, considerably higher than that reported for the lakes of the North West Territories (Various authors 1947, Miller and Kennedy 1948).

It is rather disappointing that in spite of the large series of scales, no clearcut growth trend within the year has been evident. This may perhaps be because there are fewer samples for the months of May and

FIGURE 6.



THE AGE-WEIGHT RELATIONSHIP OF OPEONGO LAKE TROUT BASED ON SCALE SAMPLES TAKEN FROM ANGLERS' CATCHES 1937 TO 1945.

June which are possibly the months of growth. There is also the possibility that selection by the fishery operates to take the fish of the entering classes as they reach a certain size. This would tend to suppress any evidence of seasonal growth since the more rapidly growing ones would be taken earliest in the season and those growing more slowly later on

Growth in Weight

The mean age weight relationship is shown in figure 6. The mean weight for a given age was derived by referring the appropriate weight

for each length class in table 8 to the length distributions and taking the means.

The Opeongo lake trout grow slowly in weight for an extended period of years. The increase is almost linear from year II to year IX with an average increment of six ounces per year. About year IX the rate increases, the increment being greatest between X and XI but continuing to be rapid up to age XVII, the age of the oldest fish captured. The course of the growth curve for the Opeongo trout is correlated with the size of the food they take. The younger age groups feed on small perch and whitefish, the older fish eat larger whitefish and large suckers. The period of slow growth occurs during the years when there is a transition in feeding habits. The growth curve as presented is biased of course by selection due to the fishing method which probably takes more of the faster growing members of the younger groups and tends to flatten the lower portion of the curve.

The average weight attained at each age and the increments between the ages are given in table 9. These mean data have been used in all calculations here that relate to weight.

TABLE 9
AGE-SIZE RELATIONSHIP OF OPEONGO LAKE TROUT

Age Group	Average length (ins)	Average weight (ozs)	Weight Increment (ozs)	Number of specimens
III	11.5	9.9	—	46
IV	13.2	15.7	5.8	143
V	14.6	21.8	6.1	268
VI	16.4	29.3	7.5	294
VII	17.9	36.3	7.0	1042
VIII	19.0	42.5	6.2	1135
IX	20.1	49.5	7.0	631
X	21.6	64.6	15.1	244
XI	23.3	87.0	22.4	86
XII	25.8	129	42	51
XIII	27.5	157	28	24
XIV	30.3	182	25	27
XV	30.5	211	29	8
XVI	29.6	200	24*	12
XVII	31.4	249	24*	7

*; increment XVI-XVII.

Minimum Production of Lake Trout in Lake Opeongo

Estimates of the minimum increment of weight of the lake trout population have been made for the years 1936, 1937 and 1938 by combining the growth data with the estimates of virtual populations. These calculations have been made as follows.

For each age group the size of the virtual population at the beginning of the fishing season following the year for which the calculation was desired, was taken from table 3. Thus if the production were being calculated for 1936 the virtual population for 1937 would be taken. All these fish would have been present in the lake throughout the previous year and their increment in weight in that season represents a production of lake trout in that season. In addition the individuals captured during the season in question also make some growth in the current year before they are taken.

Since the seasonal distribution of growth of the Opeongo lake trout is not known, a direct calculation of the amount of growth made in the current season is not possible. However, the peak of the Opeongo fishery is in July, about one-third of the total catch being taken in this month. Another third are taken in August and September. Thus it may be expected, if the seasonal growth pattern of the lake trout is similar to that of other salmonids, for example the whitefish (Kennedy 1943), the current season's growth of over half the fish taken will have been completed before they are captured. Most of the remainder would have also grown to some extent. Therefore, it has been assumed that about 75% of the current year's increment would have been attained by the catch as a whole. Hence, for the purpose of estimating production, 75% of the current season's catch from table 2 have been added to the virtual population at the end of the season. This correction for the current season's growth is probably somewhat high but such an error is balanced to a certain extent by the fact that the virtual population must be smaller than the true population.

Finally, since the three youngest age groups have not been sampled and the increments for them are not known, the total increment to age III was calculated for the virtual population in that age group.

The minimum estimates made in this manner for the three years were: 1936, 6200 lbs.; 1937, 5100 lbs. and 1938, 4040 lbs. The estimated removals in these years were respectively 9400, 7450 and 3940 lbs. In 1938 the removal and the minimum production were equal and it would appear that the fishery in that year was not a drain on the somatoplasm of the trout population. In 1936 and 1937 the removal was

somewhat greater than the minimum estimates of production, but not twice as great. Since natural mortality is not known and fish living in those years which subsequently died without being accounted for in the virtual population would have added to the production at that time, it is likely that from the point of view of removal of trout flesh the drain on the fishery was not excessive in those years either.

STATISTICS OF REPRODUCTION

Introduction

The sex ratio of the lake trout in lake Opeongo is almost perfectly 50-50. The observed percentage of females in 2774 lake trout examined over the years 1937 to 1941 was 50.3%. As usual among fish the males mature slightly in advance of the females. The relation of age to maturity in the Opeongo lake trout is similar to that reported elsewhere (Surber 1933, Royce M.S. 1943). The first male lake trout mature in lake Opeongo in year IV and all appear to be mature by year VII. While there appear to be a few females which also mature in year IV these are rare, and in general the age curve for onset of maturity in females lags a year behind that for males.

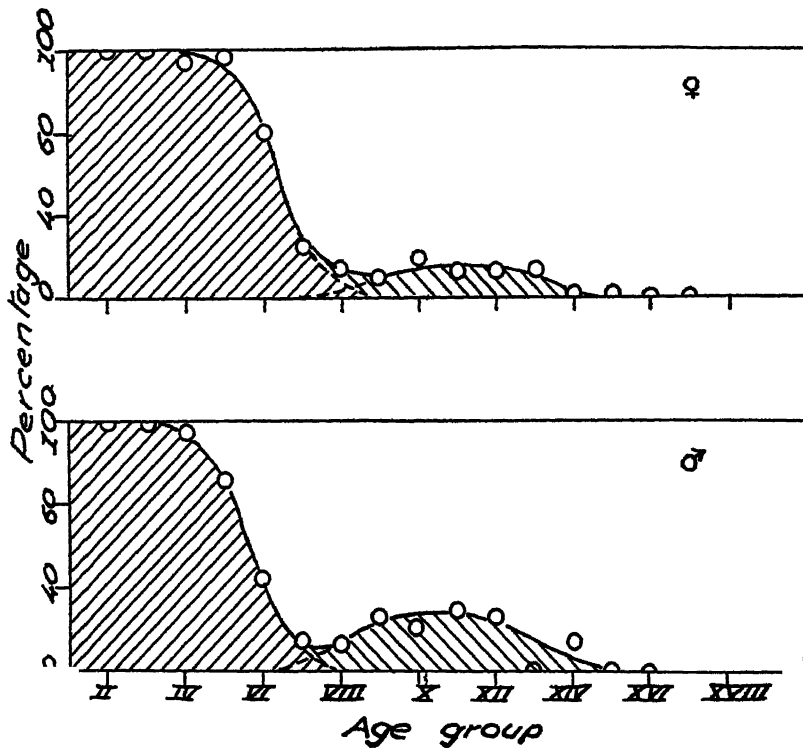
At some period after they reach maturity it is not unusual for Opeongo lake trout of both sexes to fail to produce gametes in a particular season. Such fish have been termed infertile. There appears to be no permanent damage to the gonads and it is probable that the fish spawn again in later years.

The relation between fecundity and age in Opeongo lake trout is shown in figure 7. The hatched portions show the percentage of infecund fish in each age group. Up to age VIII this hatched portion undoubtedly represents immature fish. Beyond that age it is probable that the infecund fraction of the population is made up of infertile individuals rather than of fish of delayed maturity. In the case of females this is certain since the ovary of a lake trout which has never spawned can be distinguished from that of one which has previously shed eggs.

Egg Count in Relation of Age

Estimates of the number of eggs were made on all mature females whose viscera were available and in which the eggs maturing in the current season were over 2 mm. in diameter. Except for estimates in 1937 these counts were made by measuring the diameters of 10 of the maturing eggs and weighing the ovaries. The count was determined by

FIGURE 7



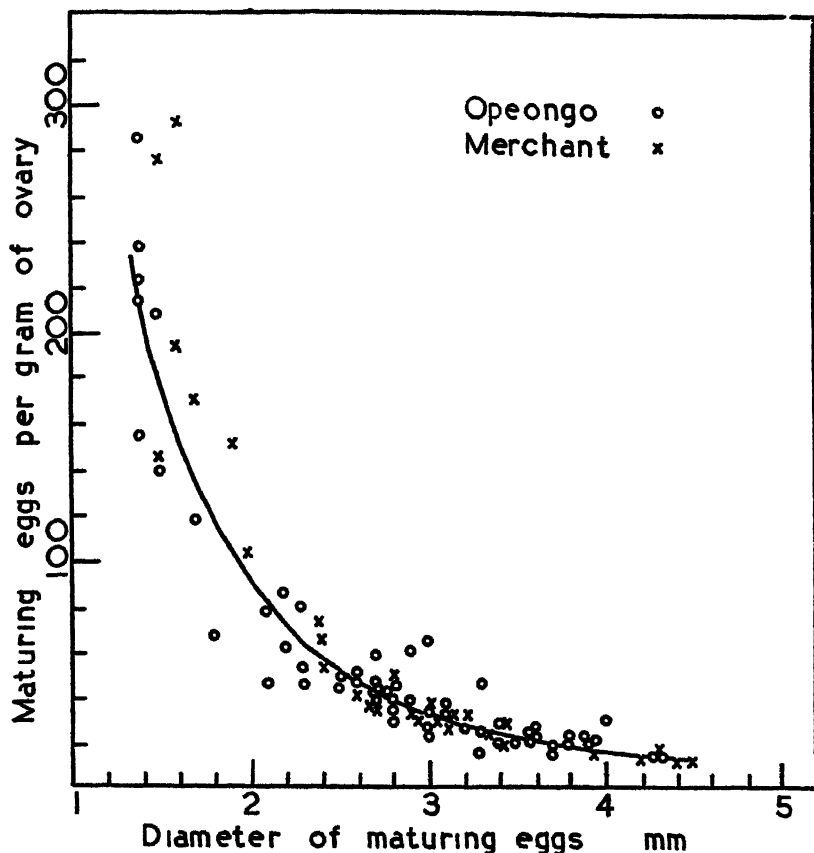
THE RELATION BETWEEN AGE AND FECUNDITY IN LAKE TROUT TAKEN BY ANGLING. THE ORDINATE IS PERCENTAGE INFECUND, AVERAGE VALUES 1937 TO 1941

using the conversion diagram given in figure 8 which it is believed holds good for lake trout in general. The egg diameters were obtained by dissecting the eggs from the stroma and measuring 10 of them placed in line together.

The average egg count in relation to age is shown in figure 9. The average number of eggs carried by an age V female is about 1200. The mean number increases along a smooth curve to 6100 at age XIII. Beyond that age the data are scanty but are grouped about an extrapolation of the same curve. The estimated mean for females of age XVII is 15,000 eggs.

A feature to be noted in figure 9 is that in a given year the egg count of fish of all ages may be consistently higher or lower than the five year mean. Thus the 1938 counts are higher and the 1943 counts are lower than the mean. This suggests that dietary conditions at the time the

FIGURE 8.

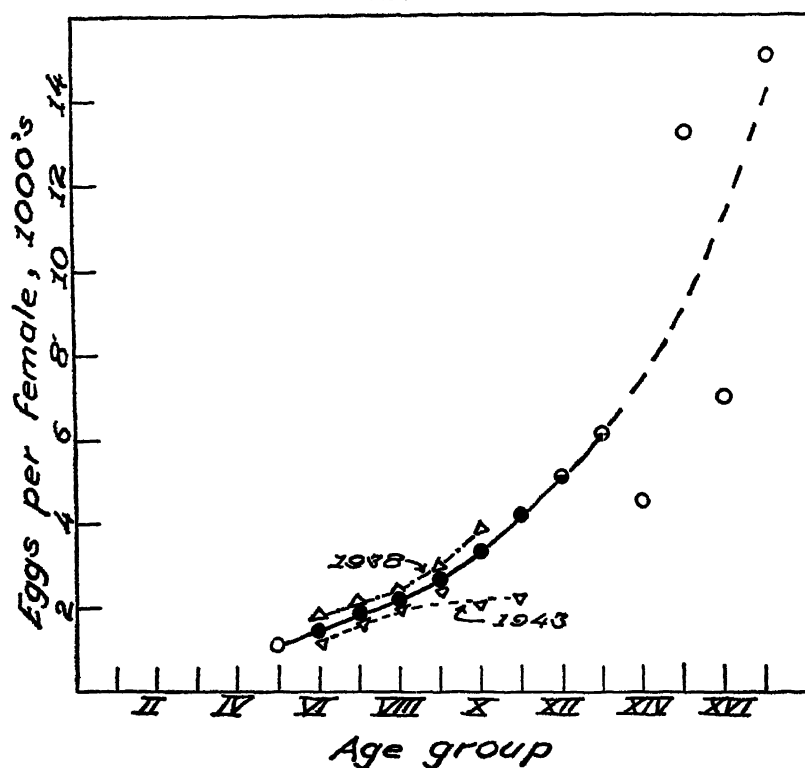


THE RELATION BETWEEN EGG DIAMETER AND NUMBER OF MATURING EGGS PER GRAM OF OVARY IN LAKE TROUT FROM OPEONGO AND MERCHANT LAKES, ALGONQUIN PARK

eggs are first laid down, probably two years previous to the one in which they mature, influence the number of eggs produced. Two seasons in which there was a marked difference in the availability of perch to the Opeongo lake trout are given in Fry (1939) figure 1.

Table 10 which gives the average number of eggs per individual for each age group supplies the final fecundity data used in calculating the spawning escapements given below. This table was constructed by combining the egg count per fecund female (figure 9) with the percentage of fecund females in each age group.

FIGURE 9



EGG COUNT IN RELATION TO AGE IN OPEONGO LAKE TROUT. THE MEAN CURVE DRAWN IS BASED ON EGG COUNTS FOR THE YEARS 1937-1941 INCLUSIVE.

TABLE 10

AVERAGE NUMBER OF MATURING EGGS PER INDIVIDUAL $m(x)$ IN LAKE TROUT OF VARIOUS AGES CAPTURED IN LAKE OPEONGO IN SUMMER BASED ON EXAMINATION OF FISH TAKEN BY ANGLING IN THE YEARS 1937 TO 1941 INCLUSIVE.

Age	Eggs per individual	Age	Eggs per individual
IV	0	XI	1750
V	58	XII	2130
VI	294	XIII	2740
VII	700	XIV	3800
VIII	1030	XV	5220
IX	1270	XVI	7840
X	1500	XVII	11900

Because of the annual fluctuation in the egg count age relationship, and probably also in the percentage infertility, the application of such mean fecundity data offers the possibility of appreciable error. However, this course has been chosen because of its statistical convenience and because of the paucity of egg count data in some seasons.

Spawning Escapement

Two partially independent estimates of the spawning escapement of the Opeongo lake trout population can be made. One of these, the *minimum spawning escapement*, has been made from the size of the virtual population and its age distribution combined with the fecundity data. The other, the *relative spawning escapement*, is based on the magnitude of the egg loss in the season preceding the spawning period and the magnitude of the fishing effort which brings about this loss of eggs. The minimum spawning escapement is an estimate of absolute magnitude whereas the relative spawning escapement can only give a ratio between two years.

Minimum Spawning Escapement. The data from which the minimum spawning escapement is calculated are contained in tables 3 and 10. Table 3 gives the virtual populations estimated to be present in the years of the investigation. As was discussed previously these values are the estimates of the size of the virtual population present at the beginning of each fishing season. Thus at least 1223 members of the 1933 year class were present at the beginning of the fishing season in which they attained age III (1936) and so on. To calculate the minimum spawning escapement these values for the beginning of a fishing season have to be referred back to the end of the season previous to it. Thus the 1223 lake trout of age III present at the beginning of the 1936 season would be considered as 1223 individuals of age II at the end of the 1935 season.

For the purpose of calculating the minimum spawning escapement in 1935 the virtual population of age groups VI to XVII at the beginning of the 1936 season were used. The number of fish in each age group within this range was multiplied by the average number of eggs per individual assigned to the age group immediately below it. Thus the 1665 fish of age VI estimated to be present in the virtual population was multiplied by 58, the average number of eggs for age V, and so on, this being the number of eggs these individuals could have been expected to have deposited the previous fall.

The minimum spawning escapement can be calculated for the spawning years 1935 to 1940 from the removal data available to the end of

1947. These values contain the averages for the older age classes mentioned on page 54. These estimates are given in table 11 together with estimates for the relative spawning escapement.

TABLE 11

ESTIMATES OF SPAWNING ESCAPEMENT AND YEAR CLASS STRENGTH. THE RATION FOR COMPARISON ARE ALL BASED ON TAKING THE 1937 SPAWNING AND THE RESULTING ESCAPEMENT, THE 1938 YEAR CLASS, AS UNITY. THE RELATIVE YEAR CLASS STRENGTH WAS ESTIMATED FROM THE NUMBER REMOVED WHICH REPRESENTED 80% OF THE VIRTUAL POPULATION ENTERING AGE 4. THESE VALUES WERE READ FROM FIGURE 3.

Spawning year	Min. Spawn Escapement		Egg loss millions	Rel. Spawn Escapement ratio to 1937	Relative year class strength
	eggs X 10 ⁶	ratio to 1937			
1931					1.7
1932					1.7
1933					1.6
1934					1.7
1935	3.43	1.99			1.4
1936	2.70	1.57	2.86	1.21	1.0
1937	1.71	1.00	2.67	1.00	1.0
1938	1.65	0.96	1.20	0.69	0.8
1939	1.30	0.76	1.44	0.89	
1940	1.22	0.71	1.03	0.76	
1941			0.70	0.56	
1942			0.47	0.79	
1943			0.74	0.97	
1944			1.15	1.14	
1945			2.23	1.40	
1946			1.25	0.68	
1947			0.92	0.66	

Relative Spawning Escapement. The lag of the estimate of the minimum spawning escapement behind the fishery is a serious limitation to its worth. It can be used only as a research tool to check the event long after its occurrence. This limitation does not restrict the use of the second estimate, the *relative spawning escapement*. The relative spawning escapement gives an estimate for the current season. The calculations for estimating the relative spawning escapement were kindly worked out for me some years ago by Dr. D. B. DeLury of the Ontario Research Foundation. They are based on his general formula (DeLury 1947) recently published.

The problem is essentially that of determining the size of the population at the end of a given fishing season and then making an estimate of the spawning potential of these survivors from a knowledge of the egg count. Since the actual values for the numbers dying naturally, $R(x;T)$, and for the fraction of the population taken per unit of fishing effort, $k(x;T)$, are not known for any age, an estimate of the absolute quantity of spawn cannot be attempted. However, the ratio of the spawning escapement in two years is not greatly affected by moderate variations in these constants so that the relative spawning may be calculated with some assurance.

The assumption is made that there are no changes in the population of eggs during the fishing season apart from loss in the fish captured. This assumption specifies what Ricker (1940) defines as a fishery of Type I. It is entirely justifiable from the point of view of recruitment since the number of lake trout eggs which are going to mature in a given season is fixed at least the year previous. Natural mortality has been considered zero; this is compensated in part by using k max for the force of fishing mortality.

The appropriate formula for calculating the relative spawning escapement is developed in the appendix as equation 14. The computations were carried out as follows:

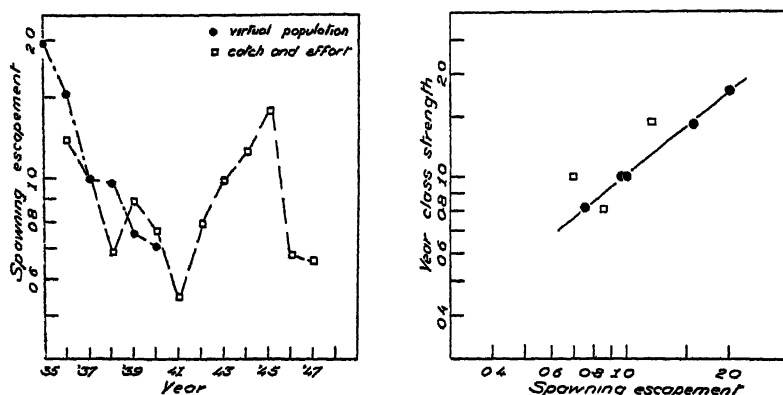
- a. A value of 0.155 was assigned to the average of $k(x;T)$; this corresponds to a level of exploitation of 30% per year over all the age groups. This value is approximately the 10 year average for the maximum level of exploitation.
- b. The total egg loss for a given year, $\sum m(x;T) K(x;T)$, was computed by summing the products of numbers of each age caught (table 2) by the mean number of eggs per individual for that age (table 10). Egg losses thus calculated are given in table 11.
- c. The appropriate values for the units of fishing effort in the years compared were taken from table 1.
- d. These numerical values were then employed in equation 14. To calculate the escapement in 1938 as compared with 1937 the values in the equation would be

$$\begin{aligned}
 S(1938, 1937) &= \frac{\text{egg loss 1938} (\exp \{-0.155 \times 22.4\} - 1)}{\text{egg loss 1937} (\exp \{-0.155 \times 16.3\} - 1)} \\
 &= \frac{1.3 \times 10^6 \times .410}{2.7 \times 10^6 \times .252} = 0.69.
 \end{aligned}$$

DISCUSSION

The trends in the minimum and the relative spawning escapements are shown in the right hand panel in figure 10. The general agreement between the estimates made by the two methods may perhaps be considered to be good, although as yet they can be compared over only four years. However, there is a reversal in the estimates for the years 1938

FIGURE 10



CHANGES IN THE ESTIMATED SPAWNING ESCAPEMENT IN THE LAKE TROUT POPULATION OF LAKE OPEONGO. THE SECOND PANEL IS THE RELATION BETWEEN THE ESTIMATED SPAWNING ESCAPEMENT AND THE ESTIMATED STRENGTH OF THE RESULTING YEAR CLASS. THE ESCAPEMENT OF 1937 AND THE HATCH OF 1938 ARE TAKEN AS UNITY. CIRCLES DENOTE ESTIMATES BASED ON THE MINIMUM SPAWNING ESCAPEMENT, SQUARES ESTIMATES BASED ON THE RELATIVE SPAWNING ESCAPEMENT.

and 1939. The estimates based on the minimum escapement show a reduction in spawning in 1939 as compared with 1938 while the estimate for the relative escapement shows the opposite. The reason for this discrepancy appears to lie in what appears to be the greatest weakness of the method by which the relative spawning escapement is derived. This weakness is the assumption that the catch per unit effort closely reflects the number of fish present. As has been pointed out earlier (page 32) there is not necessarily a close relation between the virtual population and the availability at ages VIII and younger. Unfortunately age groups V to VII make a mean contribution of about 50% to the total spawning escapement. Hence hydrological conditions which made these age groups more prone to capture in a given season would make the estimate of the relative spawning escapement considerably higher than the actual numerical strength of these age groups would

justify. The year 1940 in particular appears to have presented conditions favouring such circumstances.

On the whole, however, years of such exceptional hydrological conditions are probably rare and do not vitiate the general trend in relative spawning escapement shown in figure 10. This figure shows a steady drop in spawning escapement from 1935 to 1941, the first years following the opening of the Algonquin Park Highway. During this period spawning appears to have fallen to about one quarter of the 1935 level. With the reduced fishing intensity in lake Opeongo that was characteristic of the later war years the spawning escapement increased, rising again to approximately the 1936 level in 1945. Subsequently, increased fishing in 1946 and 1947 seems to have again reduced spawning escapement as sharply as it fell in the prewar years when access to lake Opeongo was first improved.

The possible effects of such changes in the size of the spawning escapement are not yet definitely established but unfortunately the only data at hand, that for the hatches of 1936 to 1941, indicate a strong positive correlation between year class strength and spawning escapement. This is illustrated in the second panel of figure 10. Consequently it is to be feared that the decline in spawning escapement which is the result of the increased level of exploitation of the lake trout population of lake Opeongo may materially reduce the strength of the entering year classes.

While too much reliance should not be placed on a correlation involving only five consecutive year classes, it must be pointed out that there is confirming evidence in comparative data. During this same period lake trout fishing in lake Opeongo has consistently fallen below the standard curve for the Algonquin park fisheries (Fry and Chapman 1948). This position in relation to the standard curve may be taken to indicate that the spawning in lake Opeongo may have been reduced to a level where it is limiting the size of the population.

Finally perhaps it should be pointed out that the relative spawning escapement can be estimated without any knowledge of the age of the fish concerned. Here the age-egg count relationship has been used but a size-egg count curve could be used instead. Indeed the earlier estimates for Opeongo were first made on that basis and were essentially the same as those made more recently on the basis of age composition.

NATURAL MORTALITY

The final, and indeed indispensable estimate required is that for the natural mortality occurring in the Opeongo lake trout population. A

maximum estimate of this has been attempted below by the application of DeLury's (1947) formula 5. DeLury pointed out that if the catch per unit effort bears a constant relation to the size of the population and if there is no natural mortality, then plotting the cumulative catch up to time t against the catch per unit effort at time t will result in a straight line. Deviations from such a straight line would indicate a change in the catchability of the organism or changes in the population not accounted for by the catch removed from it. Such changes in the population could result from recruitment or from natural mortality. If in constructing a curve of the type proposed by DeLury, allowance can be made for recruitment and changes in catchability, then deviations from a rectilinear relationship should give an estimate of the natural mortality involved.

Recruitment can be eliminated by considering fish of one year class only and following the removal and availability from year to year within this class.

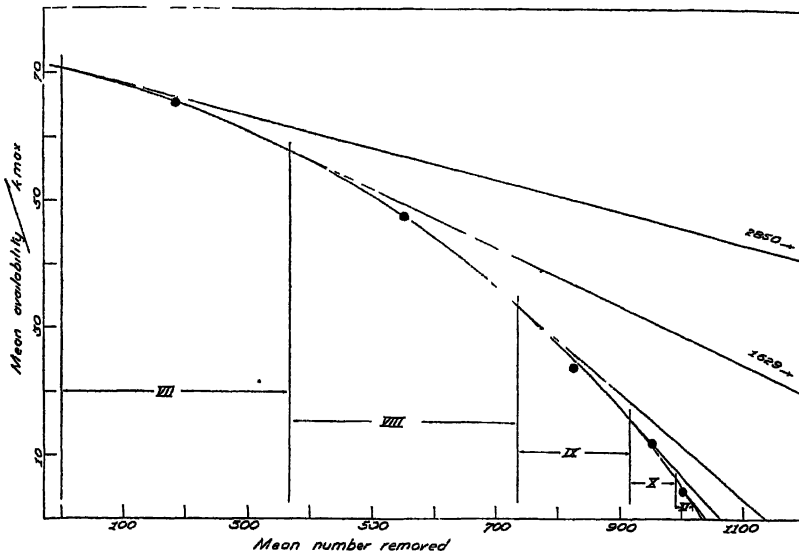
Owing to the lack of correspondence of availability with the size of the population within a fishing season the unit of time chosen has been the year. Differences between years due to annual differences in hydrological conditions can be smoothed by taking the means for each age of a series of year classes. Here year classes 1929 to 1937 have been averaged over ages VII to XI. The age groups have been restricted to this range since it is felt that the data are not yet extensive enough to warrant a consideration of all the age groups present in the fishery.

To compensate for the difference in catchability of fish of different ages the mean availability found at each age for the year classes in question has been divided by k max. Since k max is the highest possible value of the force of fishing mortality, the index of abundance derived in this manner must be a minimum value. Moreover, the younger the age the greater is the difference between k max and the true value for the force of fishing mortality. Hence the correction is more and more conservative as you proceed from the older to the younger year classes.

The curve drawn between the removal and these minimum indices of abundance is shown in figure 11. It is a curve sloping at an increasingly greater rate towards the abscissa and cutting it at a value which represents the virtual population at age VII.

If the values for the estimates of abundance had been precisely commensurate with the actual size of the population at each point considered, this curve would have been a straight line and its intercept on the abscissa would have indicated the size of the true population at

FIGURE 11.



THE CURVE USED IN MAKING THE MAXIMUM ESTIMATE OF NATURAL MORTALITY FROM AGES VII TO XI. FOR EXPLANATION SEE TEXT PAGE 58

age VII instead of the virtual one. If the curve had been constructed by using the true force of fishing mortality as a divisor, then the intercept of a tangent drawn from the beginning of age VII to the abscissa would have indicated the true population at that age. The difference between the intercept of the curve (the virtual population) and the tangent (the true population) would represent the natural mortality in age VII and subsequent years.

A tangent drawn at the beginning of age VII to the curve in figure 11 has a lesser slope than one which would be drawn to a curve corrected by the true force of fishing mortality, and hence its intercept will give a value that will be higher than the true one. Thus the intercept of this tangent with the abscissa will give a maximum estimate. Similar tangents have been drawn at the beginning of years VIII to XII and the maximum estimates for the total population at the beginning of each year which are given in table 12 were obtained graphically. The difference between any two consecutive estimates gives a maximum estimate of the total mortality for the earlier of these two years. Thus a maximum estimate for the total mortality in year VII is 1590.

TABLE 12

MAXIMUM ESTIMATES OF NATURAL MORTALITY IN THE OPEONGO LAKE TROUT
POPULATION BASED ON THE YEAR CLASSES 1929 TO 1937.

Age	Mean virt. pop.	Mean max. pop.	Ratio max./ virt.	Mean fish. mort.	Max. nat. mort.	Max. tot. mort.	Max. force nat. mort.	k min.	k max.
VII	1036	2850	2.75	369	1221	1590	0.63	0.015	0.044
VIII	667	1260	1.89	367	492	859	0.63	0.038	0.058
IX	300	401	1.34	179	72	251	0.28	0.056	0.064
X	121	140	1.16	72	28	100	0.18	0.037	0.053
XI	48	50	1.02	28	2	30	0.046	0.034	0.049

Since the catch is known (369) a maximum estimate of the natural mortality in year VII can be found as a difference between the total mortality and the catch. The maximum estimate of natural mortality in year VII is 1221.

If it is assumed that natural mortality and fishing mortality compete for the death of a given individual, then a minimum estimate for the force of fishing mortality, k min, can be calculated by substituting the required values in equations 4, 5 and 6, (page 66). At the same time a maximum value for r , the force of natural mortality, is obtained; the minimum estimate for r is, of course, zero. The estimates for r max and k min are also given in table 12. The two estimates of k give an upper and a lower limit to the mean fraction of the population of lake trout caught by the expenditure of 100 boat hours of fishing effort in lake Opeongo.

The values for the size of the population in table 12 also allow an approximate upper limit to be set to the spawning escapement, since the contribution of eggs by females of age VI and older greatly exceeds that of females of age V, the first year in which they mature. The maximum population in table 12 would have produced no more than 3.0×10^6 eggs even allowing a most generous estimate of 10,000 for the number of individuals of age V present. The virtual population given in table 12 would have produced 1.4×10^6 eggs in the previous spawning season.

The maximum estimate above was made as follows. The maximum population at the beginning of age VII was multiplied by the egg count corresponding to the age of this group the previous spawning season,

age VI. The populations for ages VIII to XI were treated in a similar manner. The twenty fish left after age XI were all arbitrarily assigned to age XIV for the purpose of determining the egg count. Finally, as mentioned above, the spawning strength at age V was set at 10,000 by extrapolation from the maximum populations estimated at the higher ages.

Fixing the maximum value for the mean spawning escapement at 3.0×10^6 places an upper limit on the average mortality rate in the first six years of life. This value is the average annual force of decrease which will reduce the value from 3.0×10^6 , the maximum value for the eggs deposited, to 2850, the maximum population at the beginning of age VII. The value of this average force is 1.15 which represents a decrement of 65% annually. Fishing mortality is included in this estimate. When it is borne in mind that there is most likely to be the highest mortality in year 0 it is not likely that the total decrement in the later years of this estimate will equal the average figure. Hence the estimate of 10,000 for the maximum population at the beginning of age VI seems to offer a comfortable margin of safety since this value allows a total mortality of 71.5% within the year.

DISCUSSION

If the ideal of the optimum catch is to so regulate the fishery that the population being exploited is maintained at a level where full advantage is taken of all of the lower stories of the food pyramid, then in all probability any management of the catch of lake trout in lake Opeongo would lead to disappointment. As far as the picture can be built up from the catch records it is clear that Opeongo contains a population of lake trout so sparse that it probably has never, during the periods for which records have been kept, been large enough to completely exploit its food supply.

Although we have no knowledge as to what conditions were like, it may be presumed from comparative data (Fry 1939^b), that in the years which preceded the exploitation of the fishery by anglers, recruitment was suppressed by the older members of the lake trout population themselves. In more recent years there are at least strong indications (figure 10) that angling can suppress recruitment by reducing the spawning stock. It can be judged from the standard curve of availability for the lake trout in Algonquin Park lakes (Fry and Chapman 1948) that the numbers of lake trout of fishable size in lake Opeongo was

somewhere in the neighbourhood of the maximum in 1936. There is no evidence that the growth rate has increased with the reduction of the population which has taken place since that time.

It would appear, therefore, that the approach to management of the Opeongo fishery should be to reinforce the recruitment by either increasing spawning escapement, supplementing spawning escapement, or reducing natural mortality at ages before the lake trout enter the fishery. Just what methods will be effective in bringing about improvement of recruitment is not known. At present the experimental planting of yearlings is being carried out. At the same time natural spawning is being studied and clues are being sought as to what are the most vulnerable periods in the life of the lake trout and what are the chief causes of early mortality.

Just how successful the attempt to describe the Opeongo lake trout population has been can be better judged by those who can view the results with greater objectivity than can the author. None of the data collected appear to be superfluous for the objective. Determination of the age composition of the catch appears to offer great advantage, particularly in dealing with the problem of recruitment.

In the case of the Opeongo lake trout population, analysis by the methods used here seems to be greatly favoured by the simplicity introduced by a low rate of natural mortality. Even moderate rates of natural mortality would make the spread between the maximum and minimum estimates too great to be of any practical value. The estimates of spawning escapement would also be far less influenced by the fishery than would be indicated by the estimates employed there. However it is to be hoped that when sufficient data are at hand to give assurance that the curve presented in figure 11 is stable that trial values of k below k_{\max} can be employed until the curve is found which gives mortalities from which a value of k can be derived which will be identical with the trial value chosen. Such a process requires first a formulation of the course of the curve between abundance and removal since the graphical method probably does not afford tangents which are sufficiently precise.

SUMMARY

1. This report is based on creel census records of the lake trout (*Cristivomer namaycush*) fishery of lake Opeongo together with data on the age, growth and fecundity of this species in that lake.
2. The various statistics summarized below have been computed from the data by means of the formulas given in the appendix.

3. Opeongo lake trout first enter the fishery at age III but are not taken in numbers until age V. Fifty percent of the fishery is drawn from ages VII and VIII. No trout over age XVII have been taken.

4. The total contribution which a year class makes to the fishery has been termed the virtual population of that year class. The virtual population in the lake at any one time is all those fish alive at that time which are destined to be captured. It is determined by summing up the contributions to the fishery from the various year classes.

5. When the catch for a given year and the size of the virtual population are known, maximum values can be calculated for the level of exploitation (percentage removed in a given year) and for the force of fishing mortality (fraction removed per unit effort).

6. The maximum value of the force of fishing mortality varies with the age of the fish. It is very low for fish of ages III and IV, reaches a maximum for age IX and probably decreases somewhat at higher ages.

7. When mean values are known for the force of fishing mortality and for the amount of fishing effort, estimates can be made of the probable fraction removed from virtual population up to the end of the current season. When this fraction reaches one half, the estimate appears stable enough to allow prediction of the total yield from the year class.

8. The yield per unit effort varies with the time of year. It is typically low in May, is highest in late June and falls again in August and early September. Fishing may recover briefly in late September. Fishing activity is negligible from October to May.

9. There is a close relation between catch per unit effort and the size of the virtual population in those age groups above year VII for which the data are adequate. The correlation is poor or non-existent at age VII and below. This lack of correlation is attributed to annual variations in catchability which appear to be more pronounced in the younger age groups.

10. The age-length and age-weight relation of the Opeongo lake trout taken by angling are described.

11. A minimum estimate of the production of trout flesh in lake Opeongo in the years 1936, 1937 and 1938 has been made from the growth curve and the size of the virtual population.

12. Lake trout in lake Opeongo mature over years IV to VII. A certain percentage of the population of mature fish fails to develop spawn in certain years. Egg counts in relation to age are presented.

13. Two estimates of the spawning escapement are presented, one based on the eggs produced by the virtual population, the other on the size of the catch and the effort required to take it.

14. These estimates indicate that the spawning escapement dropped progressively from 1935 to 1941 recovered to the 1937 level in the years from 1942 to 1945 and again decreased sharply when the fishing intensity increased again after the war.

15. For the spawning years 1935 to 1939 a positive correlation was found between the estimated spawning escapement and the resulting year class strength.

16. A maximum estimate of natural mortality can be obtained by application of DeLury's (1947) method to the relation between catch and yield per unit effort corrected by the maximum force of fishing mortality. This is presented for ages VII to XI.

17. A maximum estimate for spawning escapement was found from the maximum estimate for natural mortality and the maximum natural mortality for the years earlier than age VII deduced.

18. It is concluded that the Opeongo lake trout fishery thoroughly exploits a sparse population and that fishing mortality far outweighs natural mortality after the population has entered the fishery. There is grave danger that the level of fishing intensity has reduced the spawning escapement to a point which affects the strength of the year classes now entering the fishery. It is therefore recommended measures be taken to supplement the spawning escapement or to reduce the natural mortality in the younger year classes.

19. These data are presented as an example of the use of commonly collected biological and fishery statistics to follow a fishery and to determine its effect on the population it exploits.

APPENDIX

This mathematical analysis of the Opeongo catch statistics has been based primarily on the die-away curve following the conventional method of analyzing such statistics which on this continent has been widely employed particularly in recent years e.g. Thompson and Bell (1934) Ricker (1940, 1944, 1948) Schaeffer (1943) and DeLury (1947). In the interests of uniformity the symbols used conform to those of DeLury who appears to have published the most general formula among the groups referred to above.

Changes in the total population of trout may arise from the following sources:

- (a) An increase may come from recruitment of young fish or from immigration.

- (b) A decrease results from emigration, removal by fishing, or from death by other causes which will be termed natural.

When the ages of the captured fish are known it is possible to consider year classes separately and hence if the population is confined to a discrete body of water to eliminate from consideration all changes other than a decrease due to fishing and natural mortality.

A mathematical model of the way such a decrease takes place can be simplified by making the following assumptions:

1. The rate at which fish of an age group are caught at a given instant is proportional to:
 - (a) the numerical strength of the age group at this instant;
 - (b) the intensity of fishing effort at this instant.
2. The rate at which fish of an age group die of natural causes is proportional to the numerical strength of the age group at this instant.
3. Such characteristics as mortality rate, egg count and response to fishing depend only on the age of the fish.

The notation used to state these assumptions is taken from DeLury 1947:

x = the age of a fish. x can, therefore, have any positive value but in actual practice it will be given integral values only e.g. 0, I, II . . .

t = the time in years, for convenience the greatest value of t used is 1 year.

T = calendar year of capture.

$N(x, t; T)$ is the number of fish aged x at time t within calendar year T .

$N(x, 0; T) = N(x; T)$ is the number of fish aged x at the beginning of year T .

$e(t; T)$ = fishing intensity at time t after beginning of year T .

$E(t; T) = \int_0^t e(t; T) dt$ = total effort expended in year T up to time t .

(the unit of effort is 100 boat hours).

$E(1; T)$ will be shortened to $E(T)$.

$C(x; T)$ = Availability, number of fish aged x captured per unit effort.

$C(T) = C(x; T)$ summed over all values of x which appear in the catches.

$r(x; T)$ = force of natural mortality affecting fish aged x in year T .

$k(x; T).e(t; T)$ = force of fishing mortality affecting fish aged x at time t within year T . $k(x; T)$ is thus the fraction of the population aged x captured per unit of effort.

$m(x;T)$ = number of eggs per individual aged x in year T .

$R(x;T)$ = number of fish aged x dying of natural causes in year T .

$K(x;T)$ = number of fish aged x caught in year T .

$D(x;T) = R(x;T) + K(x;T)$ = total decrease of fish aged x in year T .

$V(x;T) = K(x;T) + K(x+1;T+1) + K(x+2;T+2) + \dots$
to the end of the table. (1)

$\frac{K(x;T) \times 100}{V(x;T)}$ = maximum percentage level of exploitation.

Assumptions 1 and 2 imply that

$$\frac{dN}{dt}(x,t;T) = -[r(x;T) + k(x;T)e(t;T)]N(x,t;T) \quad (2)$$

which integrates to give

$$N(x,t;T) = N(x;T) \exp \{-r(x;T)t - k(x;T) \cdot E(t;T)\} \quad (3)$$

it follows that

$$D(x;T) = N(x;T)[I - \exp \{-r(x;T) - k(x;T) \cdot E(T)\}] \quad (4)$$

If in addition, the force of fishing mortality bears to the force of natural mortality a reasonably constant ratio,

$$R(x;T) = \frac{r(x)}{r(x) + k(x) \cdot E(T)} \cdot D(x;T) \quad (5)$$

$$K(x;T) = \frac{k(x) E(T)}{r(x) + k(x) \cdot E(T)} \cdot D(x;T) \quad (6)$$

A special case arises when $r(x) = 0$ for all values of x . Then

$$R(x;T) = 0 \text{ for all } x \text{ and } T \quad (7)$$

which means that

$$V(x;T) = N(x;T) \text{ and} \quad (8)$$

$$N(x,t;T) = N(x;T)[\exp \{-k(x;T)E(t;T)\}] \quad (9)$$

then

$$K(x;T) = N(x;T)[1 - \exp \{-k(x;T)E(T)\}] \quad (10)$$

and

$$1 - \frac{K(x;T)}{V(x;T)} = \exp \{-k(x;T)E(T)\} \quad (11)$$

The spawning escapement ratio $S(u,v)$ between two years $T+u$ and $T+v$ will be defined by

$$S(u,v) = \frac{\sum m(x;T+u)N(x;T+u) \exp \{-k(x)E(T+u)\}}{\sum m(x;T+v)N(x;T+v) \exp \{-k(x)E(T+v)\}} \quad (12)$$

where the summation is over all values of x . If $k(x) = k$ and $m(x;T+u) = m(x)$ for all u then we have approximately

$$\frac{m(x)K(x;T+v)}{m(x)k(x;T+u)} \div \frac{m(x)N(x;T+v)[1 - \exp \{-kE(T+v)\}]}{m(x)N(x;T+u)[1 - \exp \{-kE(T+u)\}]} \quad (13)$$

(12) then becomes

$$S(u, v) = \frac{\sum m(x)K(x; T + v)[\exp \{-kE(T + u)\} - 1]}{\sum m(x)K(x; T + u)[\exp \{-kE(T + v)\} - 1]} \quad (14)$$

REFERENCES CITED

- Applegate, V. C. 1947. Growth of some lake trout *Cristivomer n. namaycush* of known age in inland Michigan lakes. *Copeia*. 237-241.
- DeLury, D. B. 1947. On the estimation of biological populations. *Biometrics* 3: 145-167.
- Fry, F. E. J. 1939^a. The position of fish and other higher animals in the economy of lakes. *Problems of lake biology* (Pub. 10 A. A. A. S.): 132-142.
- Fry, F. E. J. 1939^b. A comparative study of lake trout fisheries in Algonquin Park, Ontario. *Univ. Toronto Studies Biol. Ser.* 46, *Pub. Ont. Fish. Res. Lab.* 58: 69 pp.
- Fry, F. E. J. and V. B. Chapman. 1948. The lake trout fishery in Algonquin Park from 1936 to 1945. *Trans. Am. Fish. Soc.* 75: 19-35.
- Fry, F. E. J. and W. A. Kennedy. 1937. Report on the 1936 lake trout investigation, lake Opeongo, Ontario. *Univ. Toronto Stud.* 42, *Pub. Ont. Fish. Res. Lab.* 54: 20 pp.
- Juday, C. 1930. Growth studies of game fish in Wisconsin waters. Notes from the Biological Laboratory of the Wisconsin Geological and Natural History Survey. March 1930. Mimeographed.
- Kennedy, W. A. 1943. The whitefish, *Coregonus clupeaformis* of lake Opeongo, Algonquin Park, Ontario. *Univ. Toronto Studies Biol. Ser.* 51, *Pub. Ont. Fish. Res. Lab.* 62: 43 pp.
- Miller, R. B. and W. A. Kennedy. 1948. Observations on the lake trout of Great Bear Lake. *J. Fish. Res. Bd. Can.* 7: 176-189.
- Ricker, W. E. 1940. Relation of "Catch per unit effort" to abundance and rate of exploitation. *Fish. Res. Bd. Can.* 5: 43-70.
- Ricker, W. E. 1944. Further notes on fishing mortality and effort. *Copeia*: 23-44.
- Ricker, W. E. 1948. Methods of estimating vital statistics of fish populations. *Indiana Univ. Publ. Sci. Ser.* 15: 101 pp.
- Royce, W. E. M. S. The reproduction and studies of the life history of the lake trout *Cristivomer namaycush* (Walbaum). A thesis presented to the Faculty of the Graduate School of Cornell University for the degree of Doctor of Philosophy. 1943.
- Schaeffer, M. B. 1943. The theoretical relationship between fishing effort and mortality. *Copeia* 79-82.
- Surber, T. 1933. Rearing lake trout to maturity. *Trans. Am. Fish. Soc.* 63: 64-68.
- Thompson, W. F. and F. H. Bell. 1934. Biological statistics of the Pacific halibut fishery. *Rpt. Int'l. Fish. Comm.* 3: 47 pp.
- Various Authors. 1947. North West Canadian Fisheries surveys in 1944-1945. *Bull. 72 Fish. Res. Bd. Canada*: 94 pp.

QUERIES

QUERY: I read the paper by Anderson and Manning in the **64** September 1948 *Biometrics* and tried a simplified method of analysis as follows:

(1) At each location take the sum of the yields of the early planting minus the sum of the yields of the late planting. Consider this difference, D , to be located at the date of the intermediate planting.

(2) Plot the results, and observe that a linear regression seems adequate. Fit a linear regression, with the result

$$D = 4.04 + 0.72(t - \bar{t}), \quad s^2 = 1.89 \text{ on } 6 \text{ d.f.}$$

Hence the coefficients are 4.04 ± 0.49 and 0.72 ± 0.22 .

If Anderson and Manning's formula for regression of yield on time is accepted, then we have

$$D = y_{t(t-1)} - y_{t(t+1)} = 8b + 16c(t, - \bar{t}) = 4.04 + 0.76(t - \bar{t}).$$

Thus the simple method yields the same estimate of the first coefficient and a nearby estimate of the second coefficient. The ratio of coefficients to their estimated standard errors is 8.24 and 3.27 for the simple method as against 6.73 and 2.92 for the more complex method.

I should like to know the answers to the following questions:

- (1) Am I right in supposing that the simple method has proved more accurate in this case?
- (2) How would I expect the accuracies to compare in general?
- (3) Is the following picture correct?— The three plantings at one locality provide: a mean yield which is wholly confounded with locality; an average slope, which depends *only* on the early and late plantings, and which is confounded with the linear effect of *Lygus*; a quadratic effect, which produces an estimate of the quadratic regression coefficient of low precision. These three results may be analyzed separately. Thus the best estimate of the quadratic regression coefficient is a weighted combination of this latter estimate, and the estimate, already found above, which is based on the slope of the "observed slope" vs. "mean time of planting" plot. (The quadratic effects in each location give 0.205 ± 0.101 for the coefficient, as compared with 0.0450 ± 0.0138 from the plot. The weighted mean is 0.0479 ± 0.0135).
- (4) Why was it worthwhile to make the complicated analysis?

I should indicate, before answering the specific questions,
ANSWER: that it can be shown that the estimate of b is exactly the same using equation (20) in the Biometrics article and using that of the querist. This can be proven by expressing the value of b given in Table 7 in terms of the locality-date totals of Table 1. Similarly the equation for c given in Table 7 is almost the same as that given by the querist. In fact from Table 7

$$\frac{4096c}{3} = \frac{2}{3} [(11y_{11} - y_{21} - 10y_{31}) + (8y_{22} - y_{32} - 7y_{42}) \\ + (5y_{33} - y_{43} - 4y_{53}) + \dots].$$

As compared to querist's,

$$1344c = [7(y_{11} - y_{31}) + 5(y_{22} - y_{42}) + 3(y_{33} - y_{53}) + \dots].$$

Now let me answer the questions in the order given by the querist.

- (1) The simple method cannot possibly be more accurate, since it never utilizes more information than does our equation (20). The accuracy of the estimate of b will be exactly the same in the two cases, while that of c will be slightly greater using our equation. The reason for querist's apparently greater accuracy is that he has used a different estimate of σ^2 than that used in the article. I could have used the deviation from regression mean square [equation (28) in the article] as the error term, but since it was smaller than the experimental error I decided that the experimental error was a better estimate of the error variance for both b and c . The reason for this was that the deviation from regression mean square should equal the experimental error plus any extra component due to the regression. Hence when this deviation mean square was smaller than the experimental error, I concluded that we should use the experimental error, which was estimated with 48 degrees of freedom instead of the 7 degrees of freedom for the deviation [see equation (28) in the article]. It should be noted that querist's deviation mean square (0.236 on a per plot basis) is less than mine. It seems unreasonable to use this very low estimate, which neglects the failure of his prediction equation to fit the yields for the middle planting date at each location.
- (2) As stated above the variance of b in both cases is $\sigma^2/64$. The variance of c using our equation (20) is $3\sigma^2/4096$, and using

querist's equation $\sigma^2/1344$. Hence the efficiency of the latter is $4032/4096 = 0.985$. This shows that the loss of information using the simpler equations is only 1.5%.

- (3) The value of c given by our equation (20) is a weighted average of the two estimates mentioned by the querist. I might mention at this stage that there seems to be a slight error in the value of C given in Table 7. Apparently $C = 2299.82$ instead of 2299.62. Hence c equals -0.04722 instead of -0.04737 .
- (4) There is little to recommend one over the other of the two analyses if the regression is truly quadratic. However, a loss of information in neglecting the middle planting dates becomes more important if a higher degree equation is required. A preliminary examination using our equation (38) shows the relative efficiencies for b , c , d , and e are about .96, .84, .92 and .80 respectively. Also I doubt if there is actually much more work in computing the estimates of the constants in our equations once the solutions given at the bottom of Table 7 or the inverse matrices (40) and (43) are known. Hence if future experiments were run, these solutions could be used without going through the preliminary matrix inversions. And even though the middle dates are not used for estimation purposes, they should be used in determining the goodness-of-fit.

Finally, I should mention that one of the main purposes of writing this article was to obtain suggestions for an improved design in which the range of planting dates is no more than four weeks at a given location. The experiments analyzed here were set up before an analysis had been devised, but we hope that something better can be offered for the future. As indicated on pages 194-195, I tried several other 3-date designs but could find none to be superior to the one which had been used. It is here that the querist missed the most important part of his idea, namely that we use our same design but omit the middle date at each location. The extra plots could be used to make six replications at each location, thus increasing the efficiency of the experiment by almost 50% if the quadratic equation were adequate

$$\left[\sigma^2(b) = \frac{\sigma^2}{96}, \quad \sigma^2(c) = \frac{\sigma^2}{2016} \right].$$

If it were possible to find four more locations so that four replications were used at each of 12 locations, $\sigma^2(b)$ is the same but $\sigma^2(c) = \sigma^2/4576$. Since this involves a range of 26 weeks in planting dates, it is doubtful

if a wider range of planting dates should be used. However, if it were possible to separate the two plantings at each location by six weeks, using 8 locations and 11 planting dates, we could reduce $\sigma^2(b)$ to $\sigma^2/216$ and $\sigma^2(c)$ to $\sigma^2/4536$.

R. L. ANDERSON

65 **QUERY:** This laboratory has just completed an experiment comparing the effect of three levels of dilution of bull semen upon breeding efficiency, as measured by non-returns to first and second services within 60 to 90 days after service. In the experiment the data from one bull comprised a Latin square. Three collections (each made on a given day and consisting of the semen from two or more ejaculates combined) were made from each bull and each collection

TABLE 1
SPERM PER CC. (100,000) AND NON-RETURN PERCENTAGES

		Dilution, 1:100			Dilution, 1:150			Dilution, 1:200		
Bull	Collection			Non-			Non-			Non-
		Group	Sperm	return	Group	Sperm	return	Group	Sperm	return
1	1	2	74	57	1	50	63	3	37	46
	2	1	106	49	3	71	64	2	53	59
	3	3	125	61	2	83	67	1	63	61
2	1	3	66	66	2	75	60	1	57	65
	2	2	60	60	1	71	76	3	53	49
	3	1	75	75	3	104	67	2	78	69
3	1	1	69	69	3	107	72	2	80	78
	2	3	76	76	2	105	70	1	79	60
	3	2	72	72	1	100	63	3	75	71
4	1	1	66	66	2	66	63	3	49	62
	2	2	67	67	3	119	68	1	90	67
	3	3	74	74	1	66	77	2	49	72
5	1	3	65	65	1	85	69	2	64	62
	2	1	68	68	2	83	60	3	62	63
	3	2	74	74	3	95	76	1	71	75
6	1	2	122	71	3	81	71	1	61	66
	2	3	119	62	1	79	70	2	60	70
	3	1	140	78	2	93	69	3	70	65

was divided into three parts, with one part being diluted 1:100, one part 1:150, and one 1:200. Each third was shipped out to one third of the inseminators. The inseminators were divided into three groups as equal as possible in respect to breeding efficiency and cows bred. At time of collection a count of spermatozoan numbers was made of an aliquot undiluted sample from each collection. The numbers of sperm in the diluted samples were determined by calculation. Semen from six bulls has been used in the experiment.

I have attempted to analyze the data by analysis of variance and covariance. The table for covariance has been calculated using logarithms of sperm numbers, for the variance of actual sperm numbers was proportional to the means.

Upon examination of Table 2 you will please note that the error for Sx^2 (logs of sperm per cc.) is a negative number. The error term for Y also appears to be excessively small. Am I attempting to remove some sources of variance which are not justifiable? I would suspect that I should not be attempting to remove variance by means of all of the interactions listed in the summary tables.

The two items of information in which I am especially interested are dilution differences and bull \times dilution interaction. What should be used as error term for testing their mean squares?

TABLE 2
COVARIANCE—LOGARITHMS OF SPERM/C.C. (X) AND NON-RETURNS (Y)

Source	D.F.	Sx^2	Sxy	Sy^2
Total	53	1 3130	23.00	2,653.20
Bulls (Squares)	5	0.1777	10.16	815.42
Collections	12	0.2936	5.99	673.78
Groups	2	0.0003	0 10	42.48
Dilutions	2	0.8370	7 01	128.70
Bulls \times Dilutions	10	0.0016	-0.08	183.75
Bulls \times Groups	10	0.0013	-0 11	468.63
Groups \times Dilutions	4	0 0719	2 15	308.85
Error	8	-0 0704	-2 22	31.59

ANSWER: The calculational difficulty is this: Groups \times Dilutions and Error are not orthogonal; hence, their sums of squares are not additive. Your assumption that they are additive gives rise to the negative remainder. The correct sum of squares for

Error is the sum of the last two in the table, the degrees of freedom being $4 + 8 = 12$. This will seem reasonable if you think of the 2 degrees of freedom for error in each of the 6 squares (bulls), 12 in the pooled sum of squares.

I don't see why you have transformed the sperm numbers to logarithms. The distribution of the independent variable is not important. On the contrary, the distribution of the dependent variable affects the distribution of F : transformation to angles is worth considering. However, unless the numbers of cows in the cells of your table are very small, the transformation will not likely change decisions because your percentages lie near the middle of the range and do not vary greatly.

Since the sperm numbers for the two higher dilutions are calculated rather than observed, their use as a covariate should be avoided. The sperm numbers for these two dilutions contain no information not already used in the first dilution so that the calculations would have to be modified. Not only so, but I suspect that the dilution numbers are measured with considerably more error than are the dilution ratios. This raises a question as to their value as a covariate—you might gain nothing by the covariance analysis even if you had made independent observations of sperm numbers in every cell of the table; unless, indeed, you had replicated the observations sufficiently to get rather accurate determinations.

In your design you did not use all combinations of bulls, collections, dilutions and groups. If all were present the number of cells in the table would be $(6)(3)(3)(3) = 108$ instead of your 54. This implies some fractional replication which complicates the analysis. A suitable linear hypothesis is this:

$$y_{ijkl} = \mu + \beta_i + \gamma_j + \delta_k + (\beta\delta)_{ik} + g_l + (\beta g)_{il} + (\delta g)_{kl} + \epsilon_{ijkl}$$

where

- μ = mean effect
- β_i = bull effect ($i = 1 \dots 6$)
- γ_j = collection effect ($j = 1 \dots 3$)
- δ_k = dilution effect ($k = 1 \dots 3$)
- g_l = group effect ($l = 1 \dots 3$)
- $\beta\delta$ = bull \times dilution interaction
- βg = bull \times group interaction
- δg = dilution \times group interaction
- ϵ = experimental error

and

$$\beta \text{ is } N(0, \sigma_\beta^2)$$

$$\gamma \text{ is } N(0, \sigma_\gamma^2)$$

$$\delta \text{ is } N(0, \sigma_\delta^2)$$

$$g \text{ is } N(0, \sigma_g^2)$$

$$\beta\delta \text{ is } N(0, \sigma_{\beta\delta}^2)$$

$$\beta g \text{ is } N(0, \sigma_{\beta g}^2)$$

$$\delta g \text{ is } N(0, \sigma_{\delta g}^2)$$

$$\epsilon \text{ is } N(0, \sigma^2)$$

The dilution \times group interaction (with 4 degrees of freedom) may be split into two parts, each with 2 degrees of freedom, and these may be designated [following Yates' notation] by $DG(I)$ and $DG(J)$. Upon examination of the data, it becomes clear that $DG(J)$ is confounded with collections for bulls 1, 2 and 3 and that $DG(I)$ is confounded with collections for bulls 4, 5 and 6. It is necessary, therefore, to estimate $DG(I)$ from bulls 1, 2 and 3 only and $DG(J)$ from bulls 4, 5 and 6 only. The following analysis of variance results from this procedure:

ANALYSIS OF VARIANCE

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expectation of Mean Square
Bulls [B]	5	815.42	163.08	
Collections within bulls	12	673.78	56.15	
Dilutions [D]	2	128.70	64.35	$\sigma^2 + 3\sigma_{\beta\delta}^2 + 6\sigma_{\delta g}^2 + 18\sigma_\delta^2$
Groups [G]	2	43.48	21.24	
$DG(I)$	2	170.29	71.47	$\sigma^2 + 3\sigma_{\delta g}^2$
$DG(J)$	2	115.59	(average)	$\sigma^2 + 3\sigma_{\beta\delta}^2$
BD	10	183.75	18.38	
BG	10	468.63	46.86	$\sigma^2 + 3\sigma_{\beta g}^2$
Remainder	8	54.56	6.82	σ^2
Total	53	2653.20		

It should be noted that the expectation of the DG mean square is

$\sigma^2 + 3\sigma_{is}^2$, and not $\sigma^2 + 6\sigma_{is}^2$, because each of $DG(I)$ & $DG(J)$ is confounded with collections for one-half of the bulls.

The two tests which querist asks about in this particular problem are those for (i) dilutions and (ii) bulls \times dilutions. The latter is easy to perform being simply:

$$F = \frac{\sigma^2 + 3\sigma_{\beta\beta}^2}{\sigma^2} = 2.69 \quad \text{with d.f.} \quad n_1 = 10, n_2 = 8,$$

which is not significant at the .05 level. The former, however, is a composite test and an approximate (though biased) test would be as follows:

$$F = \frac{\sigma^2 + 3\sigma_{\beta\beta}^2 + 6\sigma_{is}^2 + 18\sigma_i^2}{\sigma^2 + 3\sigma_{\beta\beta}^2 + 6\sigma_{is}^2} = \frac{64.35}{147.68}$$

$$= .44 \quad \text{with d.f.} \quad n_1 = 2, n_2 = 2.12,$$

where the denominator is obtained by taking the mean square for bulls \times dilutions interaction and adding to it a multiple of an estimate of the σ_{is}^2 component, this estimate being obtained in the usual way:

$$\sigma_{is}^2 = \frac{71.47 - 6.82}{3} = 21.55$$

The estimate of n_2 ; i.e., $\hat{n}_2 \cong 2.12$ was calculated using the formula provided by Satterthwaite in this journal, Vol. 2, pages 110-114, December, 1946:

$$\begin{aligned} \hat{n}_2 &= \frac{\{\sigma^2 + 3\sigma_{\beta\beta}^2 + 2(\sigma^2 + 3\sigma_{is}^2 - \sigma^2)\}^2}{\frac{\{\sigma^2 + 3\sigma_{\beta\beta}^2\}^2}{k_1} + \frac{\{2(\sigma^2 + 3\sigma_{is}^2)\}^2}{k_2} + \frac{\{-2\sigma^2\}^2}{k_3}} \\ &= \frac{(147.68)^2}{\frac{(18.38)^2}{10} + \frac{(2(71.47))^2}{2} + \frac{(-2(6.82))^2}{8}} \\ &= 2.12 \end{aligned}$$

The value of F obtained is not significant.

It might be noted that the mean square for the bulls \times groups interaction leads to a significant value of F , but there does not appear to be any reason to expect such a result in this type of experiment.

BERNARD OSTLE

ABSTRACTS

- 59** BALDWIN, ALFRED L. (Fels Research Institute for the Study of Human Development, Antioch College). **Statistical Problems in the Treatment of Case Histories.**

In the use of case history material for research in human development, there frequently arises the need to apply statistical tests for the verification of interpretations and also for the discovery of relationships which demand interpretation. One of the problems raised by such attempts is the handling of the time dimension. Statistical problems in the treatment of case histories will, in this report, be illustrated in two areas.

1. Repeated measurements of a battery of variables on a single individual have been treated by various workers as a sample of measurements from the hypothetical "population" of measurements on that individual. In the analysis of such data the existence of temporal trends may be demonstrated. In economics such trends are removed to avoid "spurious" correlations. The handling of such data in psychological research requires a careful analysis of the assumptions involved.

2. Approximate temporal simultaneity of discrete events—for example anxiety attacks and threats of loss of status—are commonly used by the clinician as evidence for supporting an interpretation which relates those two events. The problem of selecting an appropriate statistic and statistical model will be discussed.

- 60** KUBIS, JOSEPH F. (Fordham University). **Statistical and Experimental Factors in the Diagnosis of Consciously Suppressed Affective Experience.**

This is a brief presentation of the development of criteria and their statistical evaluation in the diagnosis of criminal guilt. Unlike many traditional clinical problems in which the discovery of subconsciously repressed conflicts plays the major role, this practical problem concerns itself with the detection of consciously suppressed knowledge. The situation is known but the unique person to fit the situation has to be determined.

The objectification and verification of clinical intuitions in their necessary relations to both experimental and statistical control will be discussed. At present the variation in precision of clinical diagnosis is more a function of deficient experimental control and a lack of objectification of intuitions than of inadequate statistical procedures, or the unavailability of appropriate statistical techniques.

61 KOGAN, LEONARD S. and J. MCVICKER HUNT. (Institute of Welfare Research, Community Service Society of New York).
Reliability of Judges' Ratings of Case History Material.

Using an "anchored" scale of movement, various groups have rated the movement in 38 case summaries supplied by the Family Service Department. Two major null hypotheses present themselves in studying the reliability of inter-judge ratings within and between groups: (1) do groups with varying casework experience, but trained similarly, exhibit no differences in inter-judge agreement and (2) does a single group display no difference in inter-judge agreement as a result of training?

Several approaches to the statistical solution for testing these null hypotheses have been considered:

- (1) Obtain a complete matrix of inter-judge correlations within each group, treat these correlations as scores, and apply *t*-tests.
- (2) Treat the standard deviations of each set of ratings for a given case as scores and apply *t*-tests to the arrays of the two groups.
- (3) Apply analysis of variance to each group and compare the several variance estimates from one sample with that of the other.

These and similar techniques are all questionable procedures because of the assumptions involved and more appropriate methods would be valuable.

62 RABIN, ALBERT I. (Michael Reese Hospital, Chicago).
Statistical Problems Involved In Rorschach Patterning.

The present discussion is an outgrowth of efforts made to investigate the objective Rorschach criteria for schizophrenia. Hitherto the diagnosis of schizophrenia (and other disorders) by means of the Rorschach test has been made on the basis of—

- (a) A few factors in which the disorder differs from normalcy and other conditions, with statistical significance.
- (b) The presence of certain "configurations" of those and other variables which only the "experienced Rorschacher" can detect.

In the first instance, there is an overlapping with other disorders; in the second instance, the configurations detected follow, if not a strictly intuitive, certainly an insufficiently objective or quantitative method.

In order to reduce the second method to a more scientific procedure, a cluster analysis has been undertaken. The difficulties encountered in this method will be discussed; especially—

- (a) The insufficiency of present clusters for differentiation.
- (b) Quantification of qualitative (content) data.
- (c) The importance of the "single" or of the "startling" symptom or "sign".

63 CRONBACH, LEE J. (Bureau of Research and Service, College of Education, University of Illinois). **Statistical Problems in Multi-Score Tests.**

An increasing number of clinical tests report the subject's performance in a set of scores. If n people are given k scores, the resulting data can be plotted as n points in k -space. Statistical methods are necessary to describe the features of such a distribution, to compare two distributions from different groups, and to relate the distribution of cases in one k -space (test) with their distribution in another k' -space (criterion). There are serious limitations to the methods which have been used in the past to simplify this problem. These methods include treating the scores one at a time, the signs approach for limited patterns of scores, the Vernon matching method, and the discriminant function. Characteristics of a desirable statistical system for processing current clinical data are considered. Two new procedures are described: a more complex matching technique, and a pattern tabulation technique for dealing with sets of three normalized scores simultaneously.

64 STEPHENSON, WILLIAM. (Visiting Professor of Psychology, University of Chicago). **A Statistical Approach to Typology.**

Trait-universes are first defined. Thus, a limited trait universe of 2000 traits can be listed for Jung typology; 500 terms current in Rorschach interpretations likewise comprise a trait-universe, as can 800

everyday terms which one might use instead of the Rorschachian technical language. Samples from such universes can be quantified, with a person's behavior, or his personality structure, as the variables. These, when correlated with others, provide the basis of systematic typology. It is easy in this way to represent in one correlational table all the cases described, for example, in Beck's well-known monograph "*An Introduction to the Rorschach Test*". Types appear as common (non-fractional) factors, which may, or may not, be correlated factors.

Typology is thus presented as systems of trait-universes, with those concerned with motivation on the core, so to speak, and others concerned with immediate behavior at the periphery, and with many others in between. Types can appear at all levels in the system, within definable limits of error. The statistical matters involved are examined. The concept of 'significance', used for these quantifications, is defined as having the same mean value for all personalities.

The system can adequately describe all the major theories concerning personality types; it allows unique personalities to be represented; and provides a basis for the study of the intuitions of clinical and other psychologists.

65 HORN, DANIEL. (American Cancer Society). **Intra-Individual Variation in the Study of Personality.**

The psychometric approach to the understanding of an individual is limited, not by the mathematical statistical tools available, but rather through the failure to collect data for analysis in a fashion that matches a dynamic theory of personality. Dynamics implies change and the study of change necessitates a series of measurements taken over a period of time. A dynamic relationship can be evaluated statistically only by serial observations of the behavioral acts under study. The collection of data for the study of intra-individual variation and co-variation offers a way of applying present statistical techniques to current dynamic theories of personality. Otherwise a statistical analysis of cross-sectional or inter-individual data tacitly assumes a universality of static personality traits on a level with faculty psychology of the 19th century.

66 GUETZKOW, HAROLD. (University of Michigan). **Unitizing and Categorizing Problems in Coding Qualitative Data.**

The transformation of qualitative data obtained in interviews, autobiographies, free-answer questions, projective materials, and typescripts

of group meetings into a form which renders them susceptible to quantitative treatment constitutes "coding". Coding procedures implicitly involve two operations, that of separating the qualitative material into codable units, and of establishing systems of categories which can be applied to the unitized material. Examination of particular procedures suggests generalizations about the construction of category systems and the use of unitizing operations. In addition to these generalizations, it is possible to derive reliability estimates. These estimates also aid the investigator in deciding how large a sample of his data needs to be check-coded to insure the desired level of accuracy.

67 BROWN, GEORGE W. (The Rand Corporation). Basic Principles for Construction and Application of Discriminators.

Let two populations A and B have specified multivariate distributions given by $f_A(x, y, z, \dots)$ and $f_B(x, y, z, \dots)$, respectively. To classify an individual into A or B on the basis of his observed (x, y, z, \dots) , form the likelihood ratio $\lambda = f_B(x, y, z, \dots)/f_A(x, y, z, \dots)$, choosing B if $\lambda > \lambda_0$, A if $\lambda \leq \lambda_0$. This process is optimum in the sense that no other procedure can have a smaller probability of misclassifying an " A " as a " B " without having a larger probability of misclassifying a " B " as an " A ", and conversely. When f_A and f_B are multinormal, with the same variances and covariances, this process leads to a discriminator which is equivalent to the usual discriminant function.

The above process is optimum in another sense, since (if λ is properly chosen) it minimizes the expected risk, when costs of misclassification and proportions of A 's and B 's in the population to be classified can be estimated.

If C_A is the cost of misclassifying an A , C_B is the cost of misclassifying a B , and p is the expected proportion of A 's, then λ should be taken as $(pC_A)/((1-p)C_B)$. Thus the choice of λ can, in some cases, be made on a rational basis. Examples given.

The procedure is extended to include the classification of an individual into one of several populations.

68 GARDNER, ERIC F. (Syracuse University). Some Comments on Scaling Theory with a Proposal for a New Type of Scale.

Numerous instruments have been made available to the clinician and guidance counselor to aid him in his attempt to determine the

interests, attitudes, aptitudes and achievement of his clients. To interpret the results obtained with these instruments in terms of change within and differences between individuals, scales are needed.

There have been two general approaches to the scaling problem. A direct approach, in which the scale maker constructs his instrument in such a way that the score yields a scale directly without conversion, has been used in general by makers of attitude scales. Such scales generally furnish information concerning the relative rank order among the individuals to whom it is administered. Various techniques to determine scalability have been devised. A brief review of the methods of such workers as Guttman, Likert, and Thurstone will be given.

The second or indirect approach has been used by the makers of aptitude and achievement tests. In order to measure growth, interval scales are desired. The relationship between the units used and the shape of the frequency distribution of a particular trait has been used extensively in the past to define specific trait units. The work of Thorndike, McCall, and Flanagan, in which normality of distribution has been hypothesized will be discussed.

A method for the determination of a new type of scale unit which does not necessitate the assumption of a normally distributed trait will be described. (This exposition will constitute the major portion of the paper). The value of this type of scale to the clinician and counselor will be discussed.

69 VOTAW, D. F., JR. (Yale University). Compound Symmetry Tests in the Multivariate Analysis of Medical Experiments.

If experimental quantities (e.g., % CO₂ in blood, hematocrit, etc.) are measured several times on each member of a sample (of animals, say), the experimenter may wish to test statistically whether these quantities are "stable with respect to time". It is assumed that the sample is drawn from a population having a normal multivariate distribution. The (null) hypothesis of "stability" can be interpreted as a hypothesis of "compound symmetry" in the distribution. The distribution theory of criteria for testing compound symmetry was given in a recent paper [see *Ann. Math. Stat.*, Vol. 19, pp. 447-473]. The present paper gives simple methods of computing the criteria and gives the exact cumulative distribution functions for several of the criteria (when the corresponding null hypotheses are true).

70 COLLINS, SELWYN D. (Public Health Service). Intensive and Extensive Morbidity Surveys.

Morbidity data for the whole country, or for a representative sample of it, certainly would be more useful than local surveys. But the techniques of collecting morbidity data in intensive local surveys are hard to apply simultaneously in the many localities needed for a good sample.

Sickness ranges from minor nondisabling colds to chronic incapacitation with confinement to bed or to a hospital. Because of the difficulty of defining what is to be recorded as a case, rates obtained in different surveys vary considerably.

Nondisabling cases in the United States have considerable medical care. However, data on such acute cases are not so necessary, but important chronic diseases do not keep the patient from work until the later stages. Most chronic diseases are curable only in the early stages, so chronic nondisabling cases are important. A correct diagnosis and the length of time since first contracted, as well as the duration of disability, are important in recording chronic diseases.

It is essential that the transition from intensive local surveys to extensive national surveys be not accompanied by a neglect of important kinds of data.

71 SANDERS, BARKEV S. (Federal Security Agency). Measurement of the "Memory" Factor in Morbidity Surveys.

Incidence of illness for a year was derived from illnesses of specified duration current on the day of enumeration by first differencing. Completed illnesses reported in the survey as compared with the estimate was grossly deficient. Only 15 per cent of illnesses disabling 7-13 days had been reported; 43 percent of illnesses lasting 30-59 days and 75 percent of those lasting 91 to 181 days. The reported illnesses of all durations represented 26 percent of the estimated total, prior to correction for seasonality. When correction was made for seasonality, the completed illnesses reported over the year represented 59 per cent of the total. The underreporting is so great even for illnesses causing disability for 2-3 months that it does not seem that it can be attributed solely to forgetfulness. All future surveys should avoid obtaining information on disabling conditions over a year's period. They should concentrate on obtaining information on disabling illness prevalent on the day of the visit, and how long these disabilities have lasted.

- C. I. BLISS and NEELY TURNER (Connecticut Agricultural Experiment Station) and D. F. VOTAW (Yale University).
72 Dosage-Mortality Curves from Counts of Survivors.

The calculation of dosage-mortality curves is described for experiments in which only the survivors can be counted at each dosage level and the number exposed to each dose must be inferred initially from the number of survivors on untreated checks. The problem has been solved with the following assumptions. (1) In the absence of treatment, the variation in the number of larvae per plot is Poisson with m = the expected number per plot. (2) On the plots receiving any given dose the number of larvae is also subject to Poisson variation with mq = the expected number, where q is the proportion expected to survive. (3) The proportion of survivors is a function of the log-dose involving the normal distribution, so that the expected probit is $Y = \alpha + \beta X$, where X is the log-dose of toxicant and Y is the deviate (+5) corresponding to the proportionate area q of the normal curve.

Maximum likelihood has been used to obtain estimates of m , α and β , beginning with m_0 = the number of survivors on the untreated check and a graphic estimate $Y = a_0 + b_0 X$. Following an initial estimate (m_1) of m , the problem is reduced to the solution of two simultaneous equations to find corrections which are added to a_0 and to b_0 to obtain a_1 and b_1 . The procedure can be carried to i approximations ($i = 0, 1, 2, \dots$) and leads to a χ^2 test of the agreement between expected and observed results and to estimated variances of a_i and b_i . The calculation is illustrated with data from a dusting experiment on the Mexican bean beetle.

- GREEN, MELVIN W. (American Pharmaceutical Association Laboratory, Washington, D. C.) and LILA F. KNUDSEN.
73 (Food and Drug Administration, Federal Security Agency, Washington, D. C.). Statistical Variations in Contents of Dry-Filled Ampuls in Current Pharmaceutical Practice.

In 1944, the Canadian government amended the Food and Drug Act to provide weight tolerances for the contents of dry-filled ampuls. In 1946 one of our Federal agencies promulgated similar standards for a limited number of medicinal agents. In neither case was the size of the sample clearly defined, nor was it clear whether the proposed tolerances should be based upon averages or individual ampuls, although the latter was strongly implied.

To test current production against these standards and to provide a more rational basis for such standards, 612 manufacturers' lots from 5 different firms and representing 7 drugs at different dosage levels were examined. Samples of 10 were taken from each lot. About half of the lots examined met the proposed standards for individuals.

The data were subjected to variance analysis and a study of the operating characteristics. When the standard deviation (corrected to include both between-lot and within-lot variation) was plotted against the labeled quantity, it was observed that this standard deviation was substantially constant beyond about 150 mg. This suggests limits in terms of per cent below 150 mg. and in terms of a definite and constant weight above 150 mg.

For illustrative purposes, three-sigma limits were then established in one case based upon this standard deviation. Operating characteristic curves were drawn by plotting the probability of acceptance against deviation from a predetermined limit. These curves were drawn so that $P_a = 50$ per cent at the limit. Examples of several such curves were shown.

- IPSEN, J. (Yale University). **A Practical Method of Estimating**
74 the Mean and Standard Deviation of Truncated Normal Distributions. (To appear in Human Biology).

Biological data often present themselves as incomplete or truncated distributions. In the total sample, a certain number of individuals have been measured, but the remainder are known only to be all larger (or all smaller) than a known value which is the inherent truncation point. A method of estimating the mean and the standard deviation from such data is given, and tables to aid the calculation are provided.

- TUKEY, JOHN W. (Princeton University). **Separating Means**
75 into Two Groups. (Submitted to Biometrics as "Comparing individual means in the analysis of variance.")

A method of determining the significance of differences between adjacent means in a group, and of testing the significance of deviation of a straggling mean from the others of a group, is given. The test of significance is approximate. There is a worked example.

- TUKEY, JOHN W. (Princeton University). **Testing Row-by-**
76 Column Tables for Non-Additivity. (Submitted to Biometrics as "One degree of freedom for non-additivity.")

A scheme of computation is given for picking out a single degree of freedom from row-by-column tables which corresponds to non-additivity. There is a worked example.

CORNFIELD, JEROME and NATHAN MANTEL. (Public Health Service). **77 Simplified Methods for Computing the Maximum Likelihood Estimate of the Dosage-Response Curve.**

The equations defining the maximum likelihood estimates of the ED_{50} and slope of the probit line can be solved only by an iterative process. A process due to Garwood is known to converge to the correct solution with fewer cycles of computation than the better known Bliss-Fisher solution. In its original form Garwood's solution involved more computation per cycle. This paper presents tables, covering the range 0(.01)10 probits, which reduce the amount of computation per cycle so that it is now possible to take advantage of the more rapid convergence of Garwood's solution.

In many cases an extension of Karber's method, which can be derived from the maximum likelihood equations by means of three simplifying assumptions, can be used to obtain provisional estimates of the constants of the probit line. In experiments involving small numbers of animals this procedure appears to provide more accurate initial approximations than the usual graphical methods. In situations in which maximum likelihood solutions are not required, it also appears that this extension can often provide acceptable estimates with a considerable reduction in computation.

A study of Karber's method extended indicates that for finite samples the method of maximum likelihood yields biased estimates of the slope of probit line. An alternative estimation procedure which has been proposed, minimum chi-square, is subject to even more serious biases. Consequently, a modification of the method of maximum likelihood to eliminate the bias in the slope estimate is suggested.

HARSHBARGER, BOYD. (Virginia Polytechnic Institute). **78 Standard Errors Applicable to Rectangular Lattice Designs and Triple Rectangular Lattice Designs.**

In the Rectangular Lattice Designs and the Triple Rectangular Lattice Designs there are more than two equations for the variances of the difference between varietal means.

This paper presents the equation for the variances of the differences

between varietal means for all possible cases and gives a simple method for assigning the right combination of the variety means to the proper formula.

In the Rectangular Lattice Designs there are, in general, four equations. In the Triple Rectangular Lattice Design there are, in general seven equations.

- 79** BERKSON, JOSEPH, M. D. (Division of Biometry and Medical Statistics, Mayo Clinic, Rochester, Minnesota). **Minimum X^2 and Maximum Likelihood Solution in Terms of a Linear Transform, with Particular Reference to Bio-Assay** (Abstract of paper submitted to the Journal of the American Statistical Association).

A solution is derived for estimating the parameters of any function $g_i = F(x_i, \alpha, \beta)$ in terms of a linear transform $Y_i = \alpha + \beta x_i$, fulfilling the criterion of maximum X^2 , similar to the maximum likelihood solution of the integrated normal curve in terms of probits previously given by Bliss and Fisher. The solution involves successive iteration with use of specified weights and working values just as does the maximum likelihood probit solution. A table is presented giving the weight and working value for the minimum X^2 solution and the maximum likelihood solution for a number of functions in common statistical use.

- 80** BERNSTEIN, MARIANNE E. (Purdue University). **Recent Changes in the Secondary Sex Ratio in the Upper Social Strata.** (In print Human Biology).

On a stratified sample of 3898 families belonging to the American and German upper social strata, statistical investigations were made as to the effect of birth order, family size, and improved living conditions on the secondary sex ratio of Man. The data represent random samples from "Who is Who in Commerce and Industry", and "Wer Ist's", and a statistical technique was devised to check their reliability. The expected number of 1- to 4-child families with sons only is

$$S = N_1 p^1 + N_2 p^2 + N_3 p^3 + N_4 p^4$$

where p is the proportion of male offspring in the sample and N_i is the number of families with i children. The sum S was used rather than single probabilities since there is a correlation between the size of a family and the sex of the older children.

The author confirms Winston's finding that (a) the sex ratio of upper class families is significantly above average and (b) as the size of the families of men married before 1910 increases, the sex ratio of the offspring approaches the sex ratio of the total population.

However, in the families of men married since 1918 any effect of family size and birth order on the sex ratio was found to have disappeared. There is a continuous rise in the sex ratio of upper class families in the last thirty years, more pronounced in large families. Today the sex ratio of both small and large upper class families in America and Germany is about 125 to 130 males for every 100 females.

Dividing the children of a group of 880 upper class German families into three groups according to whether they were born before, during, or after World War I, a steady rise in the sex ratio was found with time. The sex ratio of 1080 children born in Germany during World War I was higher than that of their prewar and lower than that of their postwar siblings.

JOSEPH BERKSON, M. D. (Division of Biometry and Medical
81 Statistics, Mayo Clinic, Rochester, Minnesota). **Are There Two
 Regressions?**

Let u and v represent the true values of linearly related measures, d and e their respective unbiased errors; $x = u + d$; $y = v + e$.

Two types of readings are distinguished: (1) an *uncontrolled* observation, (2) a *controlled* observation. An *uncontrolled* observation is made when wishing to ascertain the value of an unknown true quantity u_i , we measure the quantity. Example: the weighing of some given material with a chemical balance. A *controlled* observation is made when we wish to bring the quantity to a value u_i . Example: the weighing out of a *specified* amount in a chemical experiment.

For the uncontrolled observation

$$\bar{x} \text{ (from } n \text{ observations, with } u = u_i) \rightarrow u, \quad n \rightarrow \infty$$

For the controlled observation

$$\bar{u} \text{ (from } n \text{ observations, with } x = x_i) \rightarrow x, \quad n \rightarrow \infty$$

Two situations are contrasted: (1) there is an existent population, as for instance a bivariate normal distribution, from which samples are drawn; (2) there is no existent population, but observations are brought into being by a controlled experiment. The values of one variate, for example u (or x), are brought to specified readings u_i (or x_i) as controlled observations, and the corresponding value of the other variate is read as an uncontrolled observation y_i (or v_i). Examples: the bio-assay

experiment in which the dosages are controlled observations; an experiment to determine the electrical resistance of a circuit by adjusting volts to specified values and reading corresponding values of current, or adjusting current and reading corresponding values of volts.

In the situation of the existent population there are two regressions: the regression of v on u , and the regression of u on v . From a random sample or a sample selected on u with corresponding observation of y , the regression of v on u is estimated by minimizing the squared residual of the dependent variate y , but if u is measured with error as x the estimated regression is biased, the more so the larger the error of x . Similarly the regression of u on v can be estimated if the independent variate is measured without error, but not if v is measured with error as y .

In the situation of the controlled experiment on u , with uncontrolled observation of the corresponding values y , minimization of the squared residual of the dependent variate y gives an estimate of the underlying regression. If the experiment is controlled on v with uncontrolled observation of the corresponding values of x , minimization of the squared residual of the dependent variate x yields an estimate of the same regression, that is, *there is only one regression*. Moreover if the measure of the independent variate is read with error as x (or y), the estimated regression is not biased thereby, even though only the squared residual of the independent variate is minimized.

Certain statistical procedures and concepts derived from the model of sampling from an existent population do not apply to the situation of the controlled experiment. Examples are (a) the linear correlation coefficient, (b) test of linearity by analysis of variance, (c) judgment of "heterogeneity" of the data by the X^2 test.

REIERSOL, OLAV. (Purdue University). **The Identifiability**
82 of a Linear Relationship between Variables which are Subject to
Error.

Let x_1 and x_2 be observed variables, let y_1 and y_2 be the "true" values of these variables and let the errors v_1 and v_2 be defined by $v_i = x_i - y_i$, $i = 1, 2$.

We suppose that there exists an exact linear relation between the true variables $y_2 = \alpha + \beta y_1$, that the errors are independent of the true variables, and that the errors are independent of each other.

A parameter is said to be identifiable if it can be determined uniquely from the joint probability distribution of the observed variables. We

shall give necessary and sufficient conditions for the identifiability of the parameter β .

- I. x_1 and x_2 stochastically independent:
 β not identifiable
- II. x_1 and x_2 stochastically dependent.
 - A. y_1 not normally distributed:
 β identifiable
 - B. y_1 normally distributed.
 - (1) Neither the distribution of v_1 nor the distribution of v_2 divisible by a normal distribution:
 β identifiable.
 - (2) Either the distribution of v_1 or the distribution of v_2 divisible by a normal distribution:
 β not identifiable.

THE BIOMETRIC SOCIETY

Plans for the Second International Biometric Conference, to be held in Geneva, Switzerland, August 30-September 2, go forward. The preliminary program provides seven scientific sessions on biometry in relation to genetics, teaching and education, experimental design, its present status, industrial applications and biological assay, ending with a session of contributed papers. The Conference Committee has obtained accommodations for members from 7-9 Swiss francs and up. To insure reservations, members planning to attend the Conference are urged to send their requests to Professor Arthur Linder, 24 Avenue de Champel, Geneva, Switzerland, before May 20.

Dr. Eric C. Wood of England urges strongly that a start be made toward international standardization of biometrical nomenclature and symbolism. We quote from Dr. Wood's letter:

"Sirs,

I have thought for some time that it would be a good thing if biometricians and statisticians could arrive at some sort of international agreement about the nomenclature and symbolism they employ. This is not the place to enlarge on this topic, but I may perhaps give two examples selected from many that have occurred to me or have been mentioned by others.

"(1) The terms 'standard error' and 'standard deviation' are at the moment used interchangeably by many people. Some attempts have been made to restrict the term 'standard deviation' (and the symbol σ) to that parameter of the population which is estimated by the 'standard error' (with the symbol s) of the sample. Whether this is a good idea or not, it should be possible to make two such terms do two distinct jobs.

"(2) When the mean result of an experiment is quoted in the form 10 ± 1.0 , what is to be understood by the quantity following the plus or minus sign? It should in my opinion be the standard error attaching to the result quoted. It should not be the standard error of a single one of the observations from which the mean result was calculated, and it should certainly not be the *probable* error, a quantity which might well disappear. Others may not agree with this opinion, but the figures in question should clearly have a unique meaning.

"I am aware that there are difficulties in achieving a large measure of standardisation at present, but it might be possible for a start to be made in certain restricted fields. I therefore suggest that the subject be placed on the programme for the next International Conference, and in the meantime the Society might form a Committee to enquire whether standardisation of the nomenclature and symbolism of statistics is desirable at all, and if so, to what extent.

"I may add that I have discussed this matter with the Officers and other Committee-Members of the British Region, and while they have not seen this letter and must not be taken necessarily to agree with its wording, they are in agreement with the general principle that the matter is worth consideration.

I am, Sirs,

Yours faithfully,

ERIC C. WOOD"

As general officers of the Society for 1949 the Council has re-elected President R. A. Fisher, Treasurer J. W. Hopkins and Secretary C. I. Bliss. Because of the long time required for ballots to reach members of the Australasian Region, the election of new Council members will be reported in the next issue.

We are happy to report the formation of two new regions of the Society. The Indian Region was organized in Allahabad during the sessions of the Indian Science Congress Association in the first week of January, under the chairmanship of P. C. Mahalanobis, Vice-President of the Region. The meeting followed an active campaign to enroll new members. Professor M. Frechet, Faculté des Sciences de Paris, and Dr. D. Schwartz, Service des Recherches Biologiques du S.E.I.T.A., 2 Ave. d'Orsay, Paris 76, have been named as provisional Vice-President and Secretary-Treasurer, respectively, of the French Region. Following an enrollment of new members the first meeting of the Region was scheduled for February or March. The proposal of a joint French-Italian Region was still under consideration when this issue of *Biometrics* went to press.

The Australasian Region, still in its infancy in the last issue of this journal, is now fully organized, with thirty-seven members, as of December, 1948, from Victoria, New South Wales, Queensland, South Australia and New Zealand. Dr. E. A. Cornish of the Council for Scientific and Industrial Research at the University of Adelaide is Vice-President, and Dr. Helen N. Turner of the McMaster Animal Health Laboratory in Guelph, N.S.W., is Secretary-Treasurer. The first meeting of the Region as a whole was held in Melbourne on January 8th with 30 or more in attendance from New South Wales, South Australia, Victoria and Australian Capital Territory. The Victorian Branch of the Region has been in active operation since last summer, with program sessions in August, October and November. Bi-monthly meetings of the Victorian Branch are planned for 1949.

The Western North American Region met jointly with the Institute

of Mathematical Statistics in Seattle at the University of Washington, on November 27. Two sessions were devoted to fishery biology, with papers in the morning by W. S. Rich, J. Neyman, R. Silliman, D. Chapman and E. L. Scott, W. F. Thompson presiding, and in the afternoon by O. E. Sette, S. C. Dodd and M. E. Schaefer, F. W. Weymouth in the chair. Regional by-laws were adopted for confirmation by the Council.

The annual meeting of the Eastern North American Region was held in Cleveland, Ohio, on December 27-30, in conjunction with the Biometrics Section of the ASA, the Institute of Mathematical Statistics and the American Public Health Association. The following regional officers were named for 1949 at the business meeting on December 30: C. P. Winsor, continuing as Vice-President; Roland H. Noel, Secretary-Treasurer; Lloyd C. Miller and Paul T. Bruyere, members of the Regional Committee for the term 1949-1951. The scientific program comprised seven sessions.

December 27. *Symposium on Statistics for the Clinician—Clinical Problems*, with Joseph Zubin as chairman and papers by A. L. Baldwin, J. F. Kubis, L. S. Kogan and J. McV. Hunt, A. I. Rabin, and L. J. Cronbach. A *Panel Discussion of Medical Statistics* with J. A. Rafferty as moderator and topics presented by D. F. Votaw, Jr., N. Scrimshaw, J. Neyman and C. W. Heath, was followed immediately by a *Symposium on Statistics for the Clinician—Proposed Solutions*, with J. Zubin as chairman and papers by E. F. Gardner, D. Horn, H. Guetzkow, and G. W. Brown.

December 28. A *Round Table on Morbidity Statistics* with H. Muench as chairman included the following discussants: T. D. Woolsey, F. Moore, S. D. Collins, B. S. Sanders, and W. T. Fales. A session on *Bioassay* had H. C. Fryer as chairman and papers by C. I. Bliss, N. Turner and D. F. Votaw, Jr., O. Kempthorne, M. W. Greene and L. F. Knudsen, and J. Ipsen.

December 29. *Effects of Errors in the Independent Variate in Regression Problems* under the chairmanship of W. E. Deming offered papers by J. Berkson, O. Reiersol and J. Neyman.

The first meeting of ENAR in 1949 was a two-day conference on *The Place of Statistical Methods in Biological and Chemical Experimentation*, at the American Museum of Natural History, New York City, on January 28 and 29. The conference was sponsored jointly with the New York Academy of Sciences and the New York Metropolitan Chap-

ter of the American Statistical Association. Papers were presented by G. W. Snedecor, G. M. Cox, F. Wilcoxon, W. J. Youden, K. A. Brownlee, R. A. Harte, C. V. Winder, H. C. Batson, C. I. Bliss, L. F. Knudsen, L. C. Miller, B. J. Vox, D. Mainland, D. D. Reid, H. M. C. Luykx, and F. E. Linder.

The Western North American Region will co-sponsor a session on June 17 at 9:00 A.M. on the application of statistics to biology, at the University of California in Berkeley. This forms part of the Fifth Regional West Coast Meeting of the Institute of Mathematical Statistics.

The Biometric Society will publish its first Directory in the spring of 1949, and each member will receive a copy gratis. It will include a list of all members through April 1949, and information concerning the organized regions, their officers, activities and by-laws.

The Society has a new home in New Haven at 321 Congress Avenue, in space provided through the kindness of the Department of Public Health of Yale University. Mail for the secretary should still be addressed to Box 1106, New Haven 4, Connecticut. Mrs. Elizabeth G. Weinman has joined the Society as Executive Assistant to the Secretary. A Standard Duplicating Machine has recently been added to the office equipment and addressograph plates have been made for all members. It is hoped that this will facilitate the speedy and efficient distribution of information to members.

NEWS AND NOTES

SPECIAL SUMMER SESSION IN SURVEY RESEARCH TECHNIQUES. The Survey Research Center of the University of Michigan will hold its special summer session in Survey Research Techniques from July 18 to August 13, 1949.

The following courses will be offered: Introduction to Survey Research, Survey Research Methods, Sampling Methods in Survey Research (elementary and advanced), Mathematics of Sampling, Statistical Methods in Survey Research, Techniques of Scaling.

In addition the introductory courses will be given from June 20 to July 16. This will permit students who are attending the full eight-week summer session of the University (June 20 to August 13) to register for the introductory courses during the second four weeks.

It is expected that this special session will attract men and women employed in market research or other statistical work and university instructors and graduate students with a particular interest in this area of social science research.

All courses are offered for graduate credit and students must be admitted by the Graduate School. Inquiries should be addressed to the Survey Research Center, University of Michigan, Ann Arbor, Michigan.

STATISTICS SUMMER SESSION IN APPLIED AND MATHEMATICAL STATISTICS. The Institute of Statistics of The University of North Carolina announces a statistics summer session, June 9 to July 19, 1949 at Chapel Hill. Intensive statistical instruction will be offered for the benefit of (1) research scholars in other sciences who want a practical working knowledge of statistical theory, (2) statistical consultants in various fields, (3) those preparing to teach statistics or to develop statistical theory, and (4) students working toward a degree in applied or theoretical statistics.

The instructional staff consists of the following professors: G. W. Snedecor, for fifteen years Director of the Statistical Laboratory at Iowa State College and author of the widely used textbook "Statistical Methods"; D. J. Finney, Lecturer in the Design and Analysis of Scientific Experiment, University of Oxford, England; J. Wolfowitz, Associate Professor, Department of Mathematical Statistics, Columbia University; and three members of the staff of the Institute of Statistics, R. C. Bose, Professor, recently from Calcutta University, India; Herbert Robbins, Associate Professor, and Gertrude M. Cox, Director.

An announcement of this summer session may be secured by writing to Director, Institute of Statistics, The University of North Carolina, Box 168, Chapel Hill, North Carolina.

STATISTICS IN MEDICINE—J. P. GRAY, M.D., Parke, Davis and Company. "The need for statistical method in medicine extends as the horizons of medicine itself are extended. The use of quantitative method has invaded each special field within medicine, perhaps to the greatest extent in the broadest field . . . public health . . . in which it is indispensable.

"There are those who attempt the practice of medicine on statistical bases, but, fortunately for patients, these are few. Many physicians are aware of inadequate background on attempting analysis and interpretation of data taken from clinical records involving few or many patients. From other physicians, however, unaware of such inadequacies, come evidences thereof, in spite of qualified editors of acceptable journals, for sometimes their writings are published. One recurring example, irritating to statistician and statistically minded reader interested in specificity and accuracy of definition of classification and other details, is found in the use of the term 'mortality rate' in referring to a number expressed per centum. Perhaps usage will be successful in changing basic definitions, but how can this path be justified when specific definitions already exist? It is NOT 'more difficult' or 'more technical' to be accurate and to use the proper term for the rate referred to—'case fatality rate.' Yet this inaccuracy continues, almost unbounded, in current medical literature.

"Undergraduate students of medicine logically might be expected to be interested, necessarily if not inherently, in quantitative method; but they frequently manifest unmistakable disinterest in statistics and statistical method, per se. If the interested teacher substitutes flank movement for frontal attack, interest of advanced undergraduate or graduate students of medicine can be aroused, probably because they have had opportunity to develop appreciation of need for analysis and appraisal of groups of observations.

"Further, an appeal usually successful involves teaching which utilizes current or recent literature in which the author has demonstrated inadequacy in soundness of presentation, or in analysis and interpretation of data. In such instances, little emphasis is required to indicate the hazard to which the author has subjected himself through his inadvertent invitation to embarrassment, at least in the estimation of the critical

reader (and physicians and undergraduate students of medicine pride themselves on membership in this category!) who, applying basis tests of statistical significance, finds that data presented do not justify conclusions drawn.

"Research workers, in medicine and in the medical sciences, comprising but a relatively small group, appreciate the importance of statistical method as a basic technique, comparable to language for transmitting thoughts and ideas to others, not to discount its use in planning.

"Statistics, therefore, has taken its place as a basic requirement in adequate preparation of the student of medicine, combining logic, mathematics, and a guiding philosophy applicable to experimental, laboratory, and clinical aspects of medicine. Admittedly, statistics has its limitations and its pitfalls, but the intelligent, honest, sound use of quantitative method will save the student, the investigator, the clinician, from pitfalls even more hazardous to be encountered by the worker oblivious to or unappreciative of the method and its usefulness."

TEACHING OF STATISTICS TO PUBLIC HEALTH WORKERS—BEN FREEDMAN, M.D., Director, Training Center, Department of Health, New Orleans, Louisiana. "The attitude of medical men towards statistics is varied. Those who realize it is an instrument for understanding events have respect for biometrics; many who do not have this realization believe in the contrite dictum that statistics can be made to corroborate anything. I believe that all public health personnel should at least know the value of the statistical method even though they may not know the detailed technique in using it. I do believe that post-graduate schools in public health are overdoing the teaching of statistics to health officers. There should be several levels of statistical courses in such schools, ranging from statistical appreciation to the detailed study of statistical method.

"I will agree that the more one knows about the techniques of a discipline the more one will appreciate the discipline, but I also believe that it is possible to teach the appreciation of a subject like statistics without subjecting the half-interested health officer to the rigors of a course in technique. Of course, those who have a background in mathematics will find the so-called rigorous course quite simple. There is too much about public health administration to be learned and to be taught in a school of public health so far as the ordinary practical health officer is concerned than worrying for six months whether he is going to pass a course in mathematics.

"Let me repeat, I believe that every public health worker should know the significance of statistics as far as it is possible, but the teaching of statistics should bear more relation to the background of the individual, to the work he is going to do, and to all remaining subjects that he has to learn at the same time."

ENGLAND—**J. Moyal**, has joined the Mathematical Statistics staff at the University, Manchester. He and **Maurice Bartlett** are collaborating on a book on Stochastic Processes . . . **Maurice G. Kendall**, Statistician and Joint Assistant General Manager to the Chamber of Shipping, "The new member of the family has arrived, named James. His growth curve, which I am plotting carefully from week to week, has for the first five weeks shown the unusual feature of sloping away from the time axis instead of towards it . . . Obviously this cannot continue and the curve will have to turn itself into a logistic sooner or later; but at the moment the rate of increase is more than satisfactory." . . . **K. A. Brownlee**, formerly with The Distillers Company, Ltd., Great Burgh, Epsom, Surrey is now with E. R. Squibb and Sons, New Brunswick, New Jersey. He reported on a confounded fractionally replicated experiment in penicillin production at a conference of The Biometric Society and Section of Biology of The New York Academy of Sciences (New York, January 28).

UNITED STATES—Reports have been received that **R. E. Blaser**, Department of Agronomy, Cornell University is back at the office carrying a normal load. We hope your luck has changed, no more "hits" . . . **R. J. Borden**, Hawaiian Sugar Planter's Association, Honolulu, sends a description, a part of which can be quoted. "We have just finished our annual year-end series of meetings with sugar plantation technologists and executives. **L. D. Bayer** kept things going smoothly, and interjected enough of his well-pointed stories to keep the groups in good humor. He certainly does have the art of pulling the right story out of the bag at the right moment." Mr. Bayer is Director of the Experiment Station of the Hawaiian Sugar Planters' Association . . . **H. L. Bush**, The Great Western Sugar Company, Longmont, Colorado, states, "We are still employing lattice designs in our variety testing programs and in some other tests. The triple lattices seem to be the best adapted to our work." . . . **M. Lois Calhoun**, is now Head of the Department of Anatomy, School of Veterinary Medicine, Michigan State College, East Lansing. She has a new DeSoto and has ideas for the growth of the Department . . . **J. H. Curtiss**, Chief, National Applied Mathematics

Laboratories has appointed himself as Acting Chief of the Section, Institute for Numerical Analysis, Los Angeles. Some readers may be interested in a portion of a letter from Mr. Curtiss. "A good deal of the work in our Statistical Engineering Laboratory deals with the application of modern statistical methods to research in physics and chemistry, and not just to industrial research. This is a virgin field for Fisherian statistics." . . . **Mary L. Dodds**, Acting Head, Foods and Nutrition Department, The Pennsylvania State College, has expressed a belief that "word of mouth is necessary for establishing understanding between a biochemist and a statistician." . . . **Daniel R. Embody**, who left Cornell University in 1942 to join the United States Navy, Bureau of Ships is now Director, Embody Statistical Laboratory, Spirit Lake, Idaho. He has a consulting and calculating business . . . **Joseph F. Pechanec**, Chief, Division of Range Research, Forest Service, Portland, writes, "The thin veneer of statistics I acquired at Iowa State has become badly eroded during the past ten years by the grind of supervisory and other activities . . . Your excellent publication, "Biometrics" brings on a cold sweat. Nevertheless, I do appreciate fully the vital need for and the extreme usefulness of tools provided by statistical methods and experimental design. To that extent, at least, my statistical training has been of immeasurable value." . . . **George Snedecor**, past-president of the American Statistical Association addressed the Central Indiana Chapter of the American Statistical Association at a dinner meeting at Purdue University, Lafayette, Indiana, on Tuesday November ninth. There were about 75 persons present, many coming from Indianapolis. The subject was "On The Design of Sampling Experiments." The same afternoon he spoke to 300 students and faculty of the university on "Selected Topics in Design of Experiments and Samplings." He was a guest of **Carl F. Kossach**, director of the newly formed statistical laboratory at Purdue University.

B I O M E T R I C S

**The Biometrics Section of the
American Statistical Association**

TABLE OF CONTENTS

Comparing Individual Means in the Analysis of Variance	JOHN W. TUKEY	99
Methods of Estimating Total Runs and Escapements of Salmon	GEORGE A. ROUNSEFELL	115
The Analysis of Extinction Time Data in Bioassay .	K. MATHER	127
The General Theory of Prime-Power Lattice Designs. III. The Analysis for p^3 Varieties in Blocks of p Plots with more than 3 Replicates	WALTER T. FEDERER	144
A Relation Between the Logarithmic, Poisson, and Negative Binomial Series	M. H. QUENOUILLE	162
The Statistical Analysis of Insect Counts Based on the Negative Binomial Distribution	F. J. ANSCOMBE	165
Queries		174
The Biometric Society		180
News and Notes		183

Material for *Biometrics* should be addressed to the Chairman of the Editorial Board, Institute of Statistics, North Carolina State College, Raleigh, N. C.; and material for Queries should go to "Queries", Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.

Officers **American Statistical Association**

President: Simon Kuznets; President-Elect: S. S. Wilks; Vice-Presidents: L. S. Kellogg, H. A. Freeman, D. S. Brady; Secretary-Treasurer: M. M. Flood; Directors: C. H. Gouliden, Isador Lubin, L. J. Reed, F. F. Stephan, W. L. Thorp, L. L. Thurstone.

Officers **Biometrics Section**

Chairman: Margaret Merrell; Secretary: Max A. Woodbury; Section Committee: Joseph Berkson, Harold F. Dorn, Alton S. Householder, Jerzy Neyman and Joseph Zubin.

Editorial Board ***Biometrics***

Chairman: Gertrude M. Cox; Members: C. I. Bliss, D. J. Finney, H. C. Fryer, O. Kempthorne, A. M. Mood, Horace Norton, G. W. Snedecor and Jane Worcester.

Membership dues in the American Statistical Association are \$3.00 a year, of which \$1.00 is for a year's subscription to the quarterly *Journal*, \$.75 is for a year's subscription to *The American Statistician*. Members of the Association may subscribe to *Biometrics* at a special rate of \$2.00. Dues for associate members of the Biometrics Section are \$3.00 a year, of which \$2.00 is for a year's subscription to *Biometrics*. Single copies of *Biometrics* are \$1.00 each and annual subscriptions are \$3.00. Subscriptions and applications for membership should be sent to the American Statistical Association, 1603 K Street, N.W., Washington 6, D. C.

Entered as second-class matter, May 25, 1945, at the post office at Washington, D.C., under the Act of March 3, 1879. *Biometrics* is published four times a year—in March, June, September and December—by the American Statistical Association for its Biometrics Section. Editorial Office: 1603 K Street, N.W., Washington 6, D.C.

COMPARING INDIVIDUAL MEANS IN THE ANALYSIS OF VARIANCE*

JOHN W. TUKEY

Princeton University

The practitioner of the analysis of variance often wants to draw as many conclusions as are reasonable about the relation of the true means for individual "treatments," and a statement by the F -test (or the z -test) that they are not all alike leaves him thoroughly unsatisfied. The problem of breaking up the treatment means into distinguishable groups has not been discussed at much length, the solutions given in the various textbooks differ and, what is more important, seem solely based on intuition.

After discussing the problem on a basis combining intuition with some hard, cold facts about the distributions of certain test quantities (or "statistics") a simple and definite procedure is proposed for dividing treatments into distinguishable groups, and for determining that the treatments within some of these groups are different, although there is not enough evidence to say "which is which." The procedure is illustrated on examples.

2. DISCUSSION OF THE PROBLEM

LET US BEGIN by considering how the latest and most advanced statistical theory would approach this problem and then explain why such a solution seems impractical. To make things more precise, let us suppose as a fictitious example that seven varieties of buckwheat; A , B , C , D , E , F , and G have been tested for yield in each of 12 locations, and that our interest is in the average yield of the buckwheat varieties in a region of which the 12 locations are a respectable sample, and in years exactly like the one in which the experiment was made. We will then have a simple and straightforward analysis of variance into varieties, locations, and interaction. We shall be concerned with the seven observed variety means and with an unbiased estimate of their variance, which will be given by $1/12$ th of the interaction mean square, which is itself on 66 degrees of freedom. What can we say about the varieties under these conditions?

We will wish to say, for example, that B and F yield better than A , C , and G , which yield better than D and E . Perhaps we might wish to add that A , C and G are not alike, although we do not know which one

*Prepared in connection with research sponsored by the Office of Naval Research.

yields better. The most modern approach would require us to proceed as follows: Write down all the possible conclusions to which we might come—the one illustrated above is one of the 120,904 similar possibilities for seven “treatments.” Then for each combination of seven true mean yields we should decide how much it would “cost” us to make each of these 120,904 decisions. Making the usual assumptions about the distribution of fluctuations in yield, we would have begun to state a mathematically well-posed problem. We are unlikely to get this far in a practical problem in my lifetime! Then we find, to our horror, that there are many competing methods of decision, and that which one risks the least will depend on the true variety yields, which we will never know. The problem is not as hopeless as it sounds, for Wald has taken a large step forward, and shown that any decision method can be replaced by one derived from *a priori* probability considerations without increasing the risk under any set of true variety yields. This is a great simplification—but the mathematical complications of dealing with 120,904 functions of seven variables are still awe-inspiring. If we were able to carry through this program—to set the risks intelligently, to carry out the mathematics, and to choose wisely among the admissible decision functions—we would surely do much better than we can hope to do now, but for the present we need to adopt a simpler procedure. (Note. The case of 3 or 4 means has been attacked within the scope of Wald’s theory by Duncan [7] using a different philosophy which emphasizes conclusions about pairs of means.)

At a low and practical level, what do we wish to do? We wish to separate the varieties into distinguishable groups, as often as we can without too frequently separating varieties which should stay together. Our criterion of “not too frequently” is a rough one, and may frequently be expressed by saying “at the 5% level” or “at the 1% level.” The meaning of these words deserves a little discussion. To the writer they do not mean, “so that an entirely nonexistent effect will be called real once in twenty times, or once in a hundred times”, but rather that “with the same sort of protection against false positives that I usually have when I make tests of significance on hypotheses suggested by the results tested, successive tests of hypotheses, tests of regression on selected variables, etc.” For these reasons, working “at the 5% level” may involve the successive use of tests, each of which yields false positives five times in a hundred, but, when used together, will yield seven, eight or nine false positives in a hundred. It is such a primitive and rough standard that we wish to combine with a primitively and roughly outlined desire to detect effects which are really there. From these primitive desires we are to seek a method.

3. THE STIGMATA OF DIFFERENCE

When the real differences between variety means are large, how do we realize this fact? Three vague criteria come naturally to mind:

- (1) There is an unduly wide gap between adjacent variety means when arranged in order of size,
- (2) One variety mean struggles too much from the grand mean,
- (3) The variety means taken together are too variable.

It is these three criteria we are going to apply in order to break up an observed set of means. We need, then quantitative tests for detecting (1) excessive gaps, (2) stragglers, (3) excess variability. These must be used when the variance of an individual observed mean is not known exactly, but rather when it is estimated from some other line of an analysis of variance table. The tests which we use must therefore be Studentized tests. Exact tests for (2) and (3) are available, but for the present we shall confine ourselves to an approximate and conservative test for (1).

If there are only two variety means, the largest gap between adjacent means is the same as the absolute value of the difference of the means. If $m_1 > m_2$, and s_m^2 is the estimated variance of a single mean, then

$$\frac{m_1 - m_2}{s_m 2^{1/2}}$$

has one-half of a t -distribution and assuming normality, exceeds 2.447 only 5% of the time when the two true means are equal and s_m is based on 6 degrees of freedom. There are good reasons based on experimental sampling (Section 9) and numerical integration (Section 8) to believe that the one-sided 5%, 2%, 1% points of

$$\frac{\text{largest gap between adjacent means}}{s_m 2^{1/2}}$$

are smaller than the corresponding two-sided percentage points of t . If this is true we will be conservative to use this ratio and the two-sided percentage points of t as a test of excessive gapping. The reasons are discussed in a later section.

The exact test of

$$\frac{m_1 - \bar{m}}{s_m}$$

where m_1 is the largest mean and \bar{m} is the grand mean has been discussed for the case of normality by K. R. Nair [4] in a very recent number of *Biometrika*. Simple and satisfactory *empirical* approximation to the

upper percentage points (between 10% and 0.1%) can be obtained by treating

$$\frac{\left(\frac{m_1 - \bar{m}}{s_m}\right) - \frac{6}{5} \log_{10} k}{3\left(\frac{1}{4} + \frac{1}{n}\right)} \quad (k > 3 \text{ means})$$

or

$$\frac{\left(\frac{m_1 - \bar{m}}{s_m}\right) - \frac{1}{2}}{3\left(\frac{1}{4} + \frac{1}{n}\right)} \quad (3 \text{ means})$$

as unit normal deviates, where s_m is based on n degrees of freedom. The adequacy of this approximation—which avoids the use of multiple entry tables—is also discussed in Section 6.

The exact test of excessive spread in general will of course be the familiar F -test (or z -test).

We propose to use these tests successively, and in the following order and manner. First, apply the gap test to break up the means into one or more broad groups. Second, apply the straggler test *within* these groups to further break off stragglers within groups. Third, apply the F -test to these new subgroups to detect excess variability. It is hard to see how to find the frequency of false positives with the whole system analytically, but the writer conjectures that, if the same level, such as 5%, is used in all three tests, the frequency of false positives will be between 1.2 and 1.6 times the level used (i.e., between 6% and 8% when a 5% level is used). This is about where the frequency of false positives stands for many repeated and result-guided tests of significance now in actual practice.

4. DETAILED PROCEDURE ILLUSTRATED BY EXAMPLES

The two examples we are going to use are those discussed by Newman [5] in connection with the use of the Studentized range. The advantages of continuing with the same examples may compensate for disadvantages of lack of simplicity, and in the case of the first example, lack of appropriateness. This first example is a 6×6 Latin square with potatoes, cited by Fisher [1] in Article 36 of *The Design of Experiments*. As first presented this example is stated to be six fertilizer treatments in a Latin Square, and Newman seems to have based his example on this discussion. Later on in the book (Article 64), Fisher points

out that these treatments were a 2×3 factorial design in nitrogen and phosphorus, so that there were specific individual degrees of freedom whose analysis was planned when the experiment was designed. These were *not* 6 treatments all on an equal footing, and overall analysis is not appropriate, *but* we shall proceed to analyze them as if they were six treatments about which there is no advance information. The six means were (A) 345.0, (B) 426.5, (C) 477.8, (D) 405.2, (E) 520.2, (F) 601.8, and the estimated standard deviation of a mean was $s_m = 15.95$.

Step 1. Choose a level of significance. For this example we shall choose 5%.

Step 2. Calculate the difference which would have been significant if there were but two varieties.

The two-sided 5% point of t on 20 degrees of freedom is 2.086. For this example, then, this least significant difference is $2.086 (2^{1/2}) 15.95 = 47.0$.

Step 3. Arrange the means in order and consider any gap longer than the value found in Step 2 as a group boundary.

Arranged in order, the means are 345.0, 405.2, 426.5, 477.8, 520.2, 601.8 and the differences $405.2 - 345.0 = 60.2$, $477.8 - 426.5 = 51.3$, and $601.8 - 520.2 = 81.6$ exceed 45.7, so that we have divided the varieties into four groups: 345.0 (A) by itself, 405.2 (D) and 426.5 (B) together, 477.8 (C) and 520.2 (E) together, and 601.8 (F) by itself.

If no group contains more than two means, the process terminates. The first example having terminated, we must pass to another to illustrate the continuance of the process. Snedecor [6] gives as Example 11.28 on p. 274 (of the 4th edition) the results of a 7×7 Latin Square with potatoes. The means were (A) 341.9, (B) 363.1, (C) 360.5, (D) 360.4, (E) 379.9, (F) 386.3, (G) 387.1 and s_m on 30 degrees of freedom was 9.52. Choosing the 5% level, for which t on 30 degrees of freedom is 2.042, we find $t(2^{1/2})s_m = 27.5$. In order, the means are 341.9, 360.4, 360.6, 363.1, 379.9, 386.3, and 387.1. No difference between adjacent means exceed 27.5, so that there is only one group at the end of Step 3.

Step 4. In each group of 3 or more means find the grand mean, the most straggling mean and the difference of these two divided by s_m . Convert these ratios into approximate unit normal deviates by finding

$$\frac{\frac{m - \bar{m}}{s_m} - \frac{6}{5} \log_{10} k}{3\left(\frac{1}{4} + \frac{1}{n}\right)}, \quad (k > 3 \text{ means in the group}),$$

$$\frac{m - \bar{m}}{s_m} - \frac{1}{2} \quad (3 \text{ means in the group}).$$

$$3\left(\frac{1}{4} + \frac{1}{n}\right)$$

Separate off any straggling mean for which this is significant at the chosen two-sided significance level for the normal.

For the Snedecor example we find $\bar{m} = 368.5$, and the most straggling mean is $m = 341.9$. The ratio is $26.6/9.51 = 2.80$. Further $\log_{10} 7 = .845$ and we are to consider

$$\frac{2.80 - \frac{6}{5} .845}{2\left(\frac{1}{4} + \frac{1}{30}\right)} = \frac{60}{51} (2.80 - 1.01) = 2.10.$$

Since the two-sided 5% level for the unit normal is well known to be 1.96, we must separate 341.9 (A).

Step 5. If Step 4 changed any group, repeat the process until no further means are separated in the old groups. The means separated off from one side of a group form a subgroup. If there are any subgroups of three or more when no more means are being separated from groups, apply the same process (Steps 4 and 5) to the subgroups.

The old group in the Snedecor example now contains 6 means, and its grand mean has increased to $\bar{m} = 372.9$. The most straggling mean is 387.1 for which $(387.1 - 372.9)/9.51 = 1.49$. The approximate unit normal deviate is $60/51 (1.49 - 0.93) = 0.66$, which is far from significance. Step 5 has produced no further effect.

Step 6. Calculate the sum of squares of deviations from the group mean and the corresponding mean square for each group of or subgroup 3 or more resulting from Step 5. Using s_m^2 as the denominator, calculate the variance ratios and apply the F -test.

In the Snedecor example, we have one group of six, for which the sum of squares of deviations is 829 and the mean square 166. The denominator is $(9.51)^2 = 90.4$ and the F -ratio 1.83 on 4 and 30 degrees of freedom, which is near the 12% point. Thus there is no overall evidence of difference in yield for these six varieties.

If varieties (B) 363.1, (C) 360.6, and (D) 360.4 had been known in advance to be different as a class from varieties (E) 379.9, (F) 386.3, and (G) 387.1, it would be fair to introduce a single degree of freedom for this

comparison, giving an analysis of variance (in terms of means) like this.

	Degrees of Freedom	Mean Square
<i>BCD</i> vs <i>EFG</i>	1	79.4
Varieties within classes	4	35
Error	30	90.4

From this we could conclude that *BCD* and *EFG* were different, even at the 1% level. There is no valid basis for this particular conclusion *unless* the classes are *uniquely* known in advance of the experiment. (There are 20 ways to split six varieties into two classes of three varieties each, so that the apparent significance of the most significant split would be expected to be at a percentage level near 1/20th of the percentage level of the whole group. The actual figures are, approximately, 0.6% and 12% and their agreement with the 1-to-20 ratio is unusually close.)

In the Fisher example, the proposed procedure gave the following result: Variety *A* (315.0) is significantly lower than varieties *D* (405.2) and *B* (426.5), these in turn are significantly lower than *C* (477.8) and *E* (520.2), and in turn these are significantly lower than *F* (601.8). All significance statements are statistical, and are at the 5% level or better.

In the Snedecor example, the proposed procedure gave the following result: Variety *A* (311.9) was significantly lower than some of the varieties *C* (360.1), *D* (360.4), *B* (363.1), *E* (379.9), *F* (386.3), and *G* (387.1) at the 5% level or better, the group of 6 varieties showed no overall evidence of internal differences at the 5% level.

These conclusions should be compared with those of Newman, who used the Studentized range to conclude in the first case that even taking *ADB* and *CEF* as two groups, neither was homogeneous. This is consistent with the result of the present analysis, but far less detailed. For the Snedecor example, Newman found that if either *A* or *F* and *G* together were made a separate group, the remainder seemed homogeneous. This is again consistent, but less detailed, since the present process finds definite reason to suppose that it is *A* which is inhomogeneous. (How much stronger is the evidence we have against *A* than against *F* and *G* is another matter.)

The writer feels that the proposed procedure is direct, reasonably simple, involves no new tables, and is ready to be used in practice and thereby put to the ultimate test.

5. THE DISTRIBUTION OF THE MAXIMUM GAP

We are interested in the following problem:

"Let a sample of k values (in our case means) be drawn from a normal distribution, of which we know only an independent estimate s of its standard deviation, based on n degrees of freedom. What is the distribution of

largest gap between ordered observed values g ,

s

The methods of Hartley, reviewed in detail by Nair [4], would allow us to solve this problem for finite n if we knew the answer for infinite n , that is for the case where we know σ , the standard deviation of the normal population.

The problem of the distribution of the largest gap in a sample of k values from a unit normal distribution can easily be attacked by experimental sampling (see Section 9). The fact that the random normal deviates of Mahalanobis [3] are printed in blocks of five leads one to study $k = 5$ and $k = 10$ first. The first 1000 blocks of five in that table were used (skipping block 768, which was marked as an error in the copy available to the author).

The results are shown below:

TABLE 1
UPPER PERCENTAGE POINTS OF THE LARGEST GAP IN AN
ORDERED SAMPLE OF k FROM A UNIT NORMAL

%	$k = 2$ theory	$k = 5$ sample of 1000 cases	$k = 10$ sample of 500 cases
10	2.33	1.86	<1.50
5	2.77	2.13	1.68
2	3.20	2.19	1.95
1	3.61	2.77	2.12

The theoretical values for $k = 2^{1/2}$ are values of $l(2^{1/2})$ and are accurate, the others are as found by experimental sampling and may deviate from accuracy by perhaps 1 or 2 in the first decimal. They are sufficiently accurate, however, to indicate that the upper percentage point decreases as k increases. Thus if we use the values for $k = 2$ we will make a conservative test. This is true for $n = \infty$, and by the nature of Hartley's expansion it will continue to hold for all reasonable values of n .

TABLE 2
QUALITY OF APPROXIMATION OF PERCENTAGE POINTS FOR THE STRAGGLER TEST

Normal percentage point minus accurate percentage point	Occurs for	Cases
0.15 to 0.20	3 means, $n \leq 15$	6
0.10 to 0.15	$\left\{ \begin{array}{l} 5\%, 3 \text{ or } 4 \text{ means, } n \leq 24 \\ 1\%, 4 \text{ means, } n \leq 11 \\ 1\%, 3 \text{ means, } n \leq 30 \end{array} \right.$	33
0.05 to 0.10	$\left\{ \begin{array}{l} 5\%, 5 \text{ means, } n \leq 24 \\ 5\%, 3 \text{ or } 4 \text{ means, } n \leq 60 \\ 1\%, 4 \text{ means, } n \leq 11 \\ 1\%, 3 \text{ means, } n \leq 120 \end{array} \right.$	21
-0.05 to -0.05	otherwise	154
-0.10 to -0.05	$\left\{ \begin{array}{l} 10\%, \text{ all cases} \\ 5\%, 9 \text{ means, } n = 10, 11 \\ 1\%, 8 \text{ or } 9 \text{ means, } n = 20 \end{array} \right.$	20

The discussion in Section 2 suggests, of course, that it would be correct and wise to find accurately the percentage points of the largest gap for various values of k and then use the appropriate values of k . This is not being suggested for the present, because:

- (1) the necessary table does not exist,
- (2) it would complicate the procedure,
- (3) there are problems in choosing the appropriate value of k ,
- (4) the simpler proposed procedure has not yet been used enough to show its characteristics.

6. THE STUDENTIZED EXTREME DEVIATE

In his recent paper, Nair [3] has given the following upper percentage points for 3 to 9 samples: (A) the 10%, 5%, 2.5%, 1%, 0.5% points for $n = \infty$, (B) the 5% points for n from 10 to 20 and 24, 30, 40, 60, 120, ∞ , (C) the 1% points for the same values of n . The accuracy of our rough approximation is most easily considered by transforming them into percentage points for the approximate unit normal deviates—these are what should be used for accuracy,—and comparing these with the percentage points of the normal—these are what we propose to use. Such a comparison has the following results, (Table 2).

Thus for about two-thirds of the cases tabulated by Nair, the error is less than 0.05, and is surely negligible in practice.

In doubtful cases, a more precise approximate test may be made as follows. Let

$$w = \frac{|m - \bar{m}|}{s_m} \quad (m \text{ an extreme mean})$$

Then treat

$$\left(\frac{k}{k-1}\right)^{1/2} \left(w - \frac{10(w-1.2)}{3n}\right)$$

as a unit normal deviate and multiply the tail area by k if only one kind of straggler (high or low) could be considered, and by $2k$ otherwise. Thus if $\bar{m} = 52$, $m = 43$, $s = 4$, $k = 13$, $n = 28$

$$w = \frac{143 - 521}{2} = \frac{9}{1} = 2.25,$$

$$\left(\frac{13}{12}\right)^{1/2} \left(2.25 - \frac{10(1.05)}{3.28}\right) = 1.041(2.25 - 0.13) = 2.20$$

Now the probability of a unit normal deviate = 2.11 is 0.01390 (from any normal table, e.g. Fisher and Yates [2] Table IX where 98.610% corresponds to a probit of 7.1200). Multiplying by 11 gives 15.3% as the approximate significance, if only low means are of interest, while the level is 30.6% when either high or low means are involved.

This approximation is discussed by Nair [4] for the case $n = \infty$, where it is due to McKay. Nair shows that it is very good indeed. The effectiveness of the term in n^{-1} may be tested by calculating the true percentage points for $w - 3n^{-1}(w - 1.2)$ from Nair's tables.

TABLE 3
UPPER PERCENTAGE POINTS FOR $w - 10/3n$ ($w = 1.2$)

5% points					1% points			
n	$k = 3$	5	7	9	$k = 3$	5	7	9
10	1.75	2.06	2.24	2.35	2.21	2.57	2.73	2.85
15	1.76	2.08	2.26	2.39	2.27	2.62	2.81	2.93
20	1.76	2.08	2.27	2.39	2.25	2.62	2.82	2.92
30	1.75	2.09	2.27	2.40	2.25	2.61	2.82	2.93
∞	1.74	2.08	2.27	2.39	2.22	2.57	2.76	2.88

The errors involved in the use of the values at the bottom of the columns of Table 3 instead of those above them can hardly ever be of practical importance.

The previous approximation is recommended for routine work since it involves less computation and no changing of significance levels. Both approximations are only good for upper percentage points in the significance test range. The latter approximation should meet all practical needs.

The writer would rarely bother with the more precise approximation except possibly for the cases where the error of the rough test is between -0.10 and -0.05 . The original experimental values are likely to be somewhat non-normal with large tails. An accurate allowance for this would be hard to compute, but it would increase the accurate percentage point slightly, more for smaller n . The rough approximation tends to compensate for this fact in most cases.

7. THE DISTRIBUTION OF LONG GAPS IN A SAMPLE OF k FROM ANY POPULATION

While we could concern ourselves with the distribution of the longest gap, the next longest gap, and so on, it seems theoretically better and practically simpler to do something somewhat different. We are going to calculate the expected number of gaps longer than a length G , which we denote by p_1 . For the sort of test considered above, there is much reason to use p_1 . For p_1 is the fraction of gaps per sample which will be falsely judged significant. If it is as bad to find two false gaps in a sample as to find one false gap in each of two samples, then we should consider p_1 .

Now we shall take the definition of a gap starting at y to be that y is the left hand of the gap. If y is the left-hand end of a gap of length at least G , we have the following table of elementary probabilities:

Event	Probability
One observation must fall between y and $y + dy$	$k dF(y)$
$k - 1$ observations must fall between $-\infty$ and y or between $y + G$ and $+\infty$	$\{F(y) + 1 - F(y + G)\}^{k-1}$
Not all $k - 1$ observations can fall between $-\infty$ and y	$-(F(y))^{k-1}$

hence

$$p_1 = k \int_{-\infty}^{+\infty} \{(F(y) + 1 - F(y + G))^{k-1} - (F(y))^{k-1}\} dF(y).$$

8. THE SYMMETRICAL CASE

If the distribution of x is symmetrical about zero, we may count only the gaps with centers to the left of the origin and then double. The expression for p_1 follows from:

Event	Probability
One observation must fall between y and $y + dy$	$k dF(y)$
$k - 1$ observations must fall between $-\infty$ and y or $y + G$ and $+\infty$	$(F(y) + 1 - F(y + G))^{k-1}$
Not all $k - 1$ observations can fall between $-\infty$ and y or $-y$ and $+\infty$	$-(2F(y))^{k-1}$

Since $y \leq -\frac{1}{2}G$, and since the result is to be doubled, we have

$$p_1 = 2k \int_{-\infty}^{-\frac{1}{2}G} \{(F(y) + 1 - F(y + G))^{k-1} - (2F(y))^{k-1}\} dF(y)$$

Making the substitutions $u = F(y)$, $h(u) = F(y) + 1 - F(y + G)$, this becomes

$$p_1 = 2k \int_{-\infty}^{-\frac{1}{2}G} h^{k-1} du - \left\{ 2F\left(-\frac{G}{2}\right) \right\}^k$$

For reasonably large G , the second term is fairly small and we can get an accurate value of p_1 with a reasonable amount of labor.

As an example, let us take the unit normal distribution and $G = 2$. Since $h(u)$ is non-analytic near 1 and has a minimum at $F(-1) = .1587$, it is natural to break the integral up into parts as follows:

$$\begin{aligned} p_1 &= 2k \int_0^{.0001} h^{k-1} du + 2k \int_{.0001}^{.001} h^{k-1} du + 2k \int_{.001}^{.01} h^{k-1} du \\ &\quad + 2k \int_{.01}^{.16} h^{k-1} du - 0.0013(2k)(h(.1587))^{k-1} - (.3171)^k \end{aligned}$$

Calculating h to four decimals, applying Simpson's rule to the range from 0 to .004, and the corresponding six-panel rule to the other three

ranges yields the following results, where the terms are given in the order of the formula above:

k	+	+	+	+	-	-	p_1
2	.00151	.01168	.07875	.16781	.00165	.10074	.15736
3	.00214	.01426	.06603	.08885	.00079	.03206	.13843
4	.00270	.01553	.04227	.04224	.00032	.01014	.09228
5	.00320	.01590	.03035	.01908	.00013	.00322	.06518
6	.00363	.01567	.02635	.00835	.00005	.00102	.05293
7	.00402	.01508	.01880	.00353	.00002	.00032	.04109
8	.00436	.01426	.01342	.00154	.00001	.00010	.03347
9	.00468	.01330	.00959	.00065	.00000	.00003	.02819
10	.00493	.01230	.00691	.00029	.00000	.00000	.02443

The value for $k = 2$ can of course be calculated directly as

$$2(1 - F(2^{1/2})) = 2(.0787) = .1574$$

The results are probably accurate to 1 or 2 in the fourth place. They can be conveniently stated as in the following table:

TABLE 4
NUMBER OF GAPS LONGER THAN 2.00 EXPECTED PER 100 SAMPLES OF 1 FROM THE UNIT NORMAL

k	2	3	4	5	6	7	8	9	10
gaps 100 samples	15.74	13.84	9.23	6.52	5.29	4.11	3.35	2.82	2.44

9. RESULTS OF EXPERIMENTAL SAMPLING

The results of the experimental sampling of 1000 sets of 5 from Mahalanobis' approximation to the unit normal are given in the following table, (Table 5).

The approximate normality of $(\text{largest gap})^{1/2}$ in this sample, as indicated by the correspondence of the last two columns between the 2% points is striking. For comparison it seemed worthwhile to examine the normality of $(\text{largest gap})^{1/2}$ for $k = 2$, where the probability of a gap $\geq G$ is $2N(G/2)$, where $N(u)$ is the unit normal cumulative. This gives the following results, (Table 6).

TABLE 5
RESULTS OF EXPERIMENTAL SAMPLING. DISTRIBUTION OF LARGEST GAPS IN
1000 SAMPLES OF 5

Cell	Number	Cum.	Equiv. Norm. Dev.	$(\text{gap})^{1/2} - 1.07$
				.23
.185-.199	2	2	-2.88	(-2.70)
.200-.299	9	11	-2.29	(-2.26)
.300-.399	20	31	-1.87	-1.90
.400-.499	28	59	-1.56	-1.57
.500-.699	97	156	-1.01	-1.00
.700-.899	141	297	-0.53	-0.52
.900-1.099	172	469	-.08	-.09
1.100-1.299	149	618	0.30	0.30
1.300-1.499	126	744	0.66	0.68
1.500-1.699	110	854	1.05	1.00
1.700-1.899	56	910	1.34	1.34
1.900-2.099	36	946	1.61	1.64
2.100-2.299	24	970	1.88	1.90
2.300-2.499	11	981	2.07	2.12
2.500-2.699	8	989	2.29	(2.51)
2.700-2.899	4	993	2.46	(2.77)
2.900-3.099	4	997	2.75	(2.99)
3.100-3.299	2	999	3.09	(3.20)
...				
4.000-4.099	1	1000	∞	

Here the fit is good between the 10% points. This suggests that the $(\text{largest gap})^{1/2}$ may be a convenient interpolation variable.

The number of cases ≥ 2.00 actually found was 68, while the number to be expected according to the last section was 65.2 less an allowance for large double gaps which might amount to one unit. Finding 68 instead of 64 is a deviation of 0.5σ , and is highly reasonable.

For $k = 10$, the count was only made for gaps ≥ 1.5 , with the following results, (Table 7).

The fit here is reasonably good out to the 5% point. Since theory predicts about 12.2 beyond 2.00 instead of 9 observed, there is no serious disagreement here.

If we want to make real use of this $(\text{gap})^{1/2}$ variable, we may use the known percentages beyond 1.414, found for k between 2 and 10 in the last section to fix lines in the plane of the mean and standard deviation

TABLE 6
CUMULATIVE FOR (LARGEST GAP)^{1/2} IN SAMPLES OF 2 FROM THE UNIT NORMAL

%	gap	(gap) ^{1/2}	Equiv. Norm. Deviate	(gap) ^{1/2} - .98
				.44
1	.0177	.134	-2.33	(-1.95)
2	.0357	.189	-2.05	(-1.80)
5	.0891	.299	-1.64	(-1.55)
10	.1781	.423	-1.28	-1.26
20	.360	.600	-0.84	-0.86
50	.960	.980	0.00	0.00
80	1.825	1.353	0.84	0.85
90	2.350	1.536	1.28	1.26
95	2.794	1.672	1.64	(1.57)
98	3.308	1.821	2.05	(1.91)
99	3.650	1.914	2.33	(2.12)

TABLE 7
RESULTS OF EXPERIMENTAL SAMPLING
DISTRIBUTION OF LARGEST GAPS IN 500 SAMPLES OF 10

Cell	Number	Cumul.	Equiv. Norm. Deviate	(gap) ^{1/2} - 0.85
				.24
-1.499	454	454	1.33	1.38
1.500-1.599	15	469	1.54	1.53
1.600-1.699	9	478	1.71	1.68
1.700-1.799	9	487	1.94	1.82
1.800-1.899	2	489	2.01	1.98
1.900-1.999	2	491	2.10	2.08
2.000-2.199	1	492	2.14	(2.39)
2.200-2.399	2	494	2.26	(2.48)
2.400-2.599	3	497	2.51	(2.93)
2.600-2.799	2	499	2.88	(3.13)
...				
3.100-3.199	1	500	∞	

of the approximation. A little bold, dashing, freehand, two-dimensional interpolation produces the following results:

TABLE 8
TENTATIVE BEHAVIOR OF $(\text{LARGEST GAP})^{1/2}$ FOR SAMPLES OF k FROM THE UNIT
NORMAL

k	Parameters		Levels for $(\text{gap})^{1/2}$			Levels for gap		
	m	s	5%	2.5%	1%	5%	2.5%	1%
2	0.98	0.43	1.69	1.82	1.98	2.8	3.3	3.9
3	1.03	0.36	1.62	1.74	1.87	2.6	3.0	3.5
4	1.06	0.27	1.50	1.59	1.69	2.3	2.5	2.8
5	1.06	0.23	1.43	1.51	1.60	2.0	2.3	2.6
6	1.06	0.22	1.42	1.49	1.57	2.0	2.2	2.5
7	1.04	0.21	1.39	1.45	1.53	1.9	2.1	2.3
8	1.02	0.21	1.37	1.43	1.51	1.9	2.0	2.3
9	1.00	0.21	1.35	1.41	1.49	1.8	2.0	2.2
10	0.99	0.22	1.33	1.40	1.48	1.8	2.0	2.2

By a stroke of luck, the levels for the gap itself might be accurate to one or two tenths. These are, of course, unstudentized levels.

REFERENCES

- [1] R. A. Fisher 1935-1947, *The Design of Experiments*, 4th edition 1947, Oliver and Boyd, Edinburgh.
- [2] R. A. Fisher and F. Yates, 1938-1948, *Statistical Tables*, 4th edition 1948, Oliver and Boyd, Edinburgh.
- [3] P. C. Mahalanobis et al 1934, "Tables of Random Samples from a Normal Population", *Sankhyā* 1 (1933-34), pp. 289-328.
- [4] K. R. Nair 1948, "The distribution of the extreme deviate from the sample mean and its Studentized form". *Biometrika* 35 (1948) pp. 118-144.
- [5] D. Newman 1939, "The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation". *Biometrika* 31 (1939-40) pp. 20-30.
- [6] G. W. Snedecor 1937-1946, *Statistical methods*, 4th edition 1946, Collegiate Press, Ames, Iowa.
- [7] David B. Duncan 1947, Iowa State College Thesis, iii + 117 pp.

METHODS OF ESTIMATING TOTAL RUNS AND ESCAPEMENTS OF SALMON

GEORGE A. ROUNSEFELL

*Chief, Atlantic Salmon Investigations
Branch of Fishery Biology
Fish and Wildlife Service
U. S. Department of the Interior*

In creating the management programs necessary for the conservation of salmon fisheries, special attention should be given to salmon investigations of the past to profit by their successes and failures. Investigations of the Fraser River sockeye salmon afford excellent illustrative material for developing management procedures. This fishery was studied by several famous biologists of the past, including Edward E. Prince, Richard Rathbun, David Starr Jordan and Charles Henry Gilbert. They have been followed by Wilbert A. Clemens, Lucy S. Clemens, Henry O'Malley, Willis H. Rich, R. Earle Foerster, William E. Ricker, George A. Rounsefell, George B. Kelez and Will F. Thompson, all of whom have published on the fishery. These Fraser River investigations have been pursued over such a long period as to make their study of peculiar importance in the evaluation of methods. Having personally collected, analyzed and published statistical data on gear, catches and abundance up to 1934, I have drawn freely upon them.

THE FIRST STEP in any quantitative biological investigation is to delimit the population being studied. Salmon catches often represent mixed populations as the salmon bound for any particular river may traverse numerous waterways before reaching the river mouth, and be taken along with salmon bound for other rivers. Extensive marking and tagging experiments have partially delimited the populations for many Alaskan rivers but much remains to be done, especially for the great red-salmon streams in Bristol Bay. The Fraser River sockeye run has been so dominant in its area that this problem has not assumed as great importance although there is still insufficient knowledge as to what portion of the run enters the Gulf of Georgia from the north. Tagging experiments by Clemens showed that sockeyes using this northern route were bound chiefly for the Fraser River; and there is some

evidence that a larger proportion enter the Gulf north of Vancouver Island during the warmer years.

Once the population has been circumscribed, the problem is to determine within reasonable limits what factors may be responsible for annual variations in its size. The methods developed by Baranov (1918) and greatly improved by Ricker (1940, 1944) for the determination of the abundance of populations of marine fishes are not applicable, because any particular salmon is not subject to capture throughout the season but only during that part of the season during which it runs the gauntlet of the gear; and because the total population dies after spawning and is therefore available during but one season.

For wise management it is desirable to know within reasonable limits the actual numbers of fish both in the catch and in the escapement through the fishery toward the spawning grounds. In some of the smaller salmon rivers it has been possible to erect weirs and count the number of salmon escaping through the fishery, and in a few of the larger rivers they are counted while ascending fishways. This escapement, when added to the total catch attributed to fish spawned in the same river, gives the total run of the season. Such counts have been made since 1921 at Karluk River, Alaska, and since 1935 at Bonneville Dam, on the Columbia River, but because such weir counts have proved impractical in most larger rivers, recourse must be had to other methods. The Fish and Wildlife Service, the Fisheries Research Board of Canada, and the International Pacific Salmon Fisheries Commission, have marked and released adult salmon near the mouth of a river to determine the number of spawners from the proportionate numbers of marked and unmarked fish found on the spawning grounds. The Fish and Wildlife Service is also perfecting the use of aerial photography to cover the vast lake and river systems in Bristol Bay. However, such methods do not give us an insight into the past, and it will take many years to furnish a long enough series to show what variations to expect.

For the Fraser River, Rounsefell and Kelez (1938), developed indices of relative abundance based on the catch-per-unit of fishing effort by standardized amounts of fishing gear, but all such indices fail to show the total population and are therefore somewhat limited in their application.

Using the data given by Rounsefell and Kelez (1938) an estimate has been developed of the total salmon run of each season to the Fraser River from 1894 to 1945. This method, which may find application in other salmon investigations, has been adapted from a method developed by Professor D. B. DeLury (1947) of the Ontario Research Foundation.

The notation used by Professor DeLury is as follows:

- C = Catch per unit of fishing effort
 N = Number in population at any time (t)
 E = Number of units of fishing effort
 $b = k \log_{10} e$ in which k is a constant

He has shown that

$$(1) \log C(t) = \log (k N_0) - b E(t)$$

This is equivalent to a linear equation in which

$$(2) \log y = \log a - bx$$

By determining the equation at C_0 the total population equals $k N_0/k$ before fishing commences.

In a closed population the total population at the commencement of fishing may be determined by periodically plotting $\log C$ against the accumulated effort, E , throughout the season and then fitting a linear regression to the data in order to determine the slope, b , and the intercept, $\log (k N_0)$.

In adapting this method to the salmon data there is available for most years only the average catch per unit of gillnet effort for each year. Since the same population was not present each year it was necessary to determine the relationship between the catch per unit of fishing effort and the number of units of effort by a multiple covariance analysis, using the data of Rounsefell and Kelez for the 39 years from 1896 through 1934. Three variables were used: 1) The index of abundance from traps (expressed in logarithms). This I regard as the best measure of the runs before they reach the river. 2) The number of gillnets fished. 3) The catch per gillnet (expressed in logarithms). Figure 1 shows the regression of the logarithm of catch per gillnet on number of gillnets when the abundance is held constant. The effect of cyclic differences was removed by a covariance analysis (see Table 1).

From this regression the catch per unit of effort at 0 units is easily calculated for each season. The notation used is as follows:

- Y = log of catch per unit of gillnet effort
 X_1 = log of trap index of abundance
 X_2 = number of gillnet units of fishing effort
 C' = antilog of Y , or estimated catch per unit of gillnet effort
 (o) and (p) = number of gillnet units in X_2 at 0 nets and at any particular number of nets (p)
 \bar{x}_1 = mean of X_1
 C = observed catch per unit of gillnet effort

- $C'_{x_2}(0)(\bar{x}_1)$ = calculated catch per gillnet unit when X_2 equals 0, with X_1 at its mean \bar{x}_1
 $C'_{x_2}(p)(\bar{x}_1)$ = calculated catch per gillnet unit when X_2 equals p , with X_1 at its mean \bar{x}_1
 $C_{x_2}(p)$ = observed catch per gillnet unit when X_2 equals p for any given season with p number of nets
 $C'_{x_2}(0)$ = unknown catch per gillnet unit when X_2 equals 0, for any given season

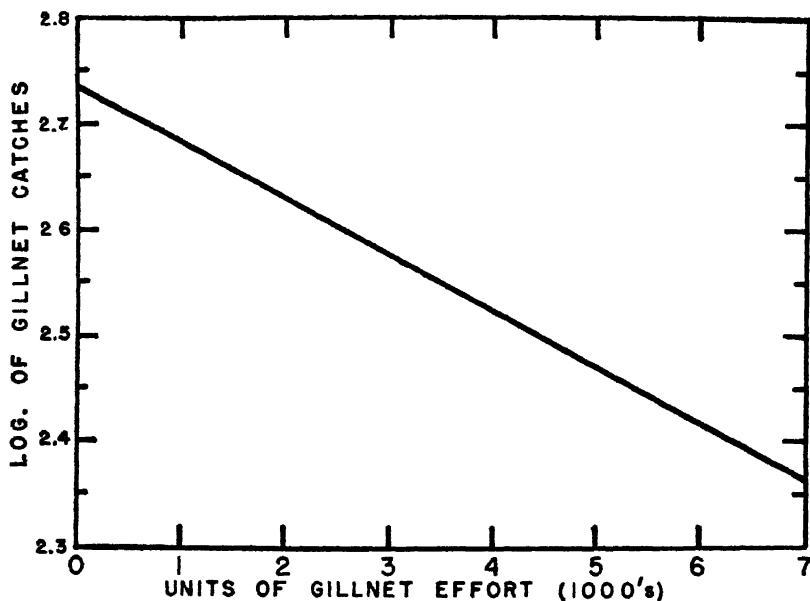


FIGURE 1

The regression of the logarithm of the catch per gillnet on number of gillnets when the abundance is held constant.

From the multiple regression formula: $\hat{Y} = 0.953100 + 0.940834 X_1 - 0.000529 X_2$, the estimate of Y with X_1 at \bar{x}_1 and X_2 at 0 is 2.733, the antilog $C'_{x_2}(0)(\bar{x}_1) = 541$. Similarly, it is easy to obtain $C'_{x_2}(p)(\bar{x}_1)$ by substituting the units of effort, p , for any given year for X_2 in the formula with X_1 at \bar{x}_1 . The observed catch per gillnet unit of effort for the season $C_{x_2}(p)$ is already known. The catch per unit of effort for the given season at 0 units of effort, $C'_{x_2}(0)$, is then easily obtained by simple proportion:

$$\frac{C'_{x_2}(0)(\bar{x}_1)}{C'_{x_2}(p)(\bar{x}_1)} = \frac{C'_{x_2}(0)}{C_{x_2}(p)}$$

TABLE 1

MULTIPLE REGRESSION OF THE LOGARITHM OF THE CATCH PER GILLNET UNIT OF EFFORT, Y , ON THE LOGARITHM OF THE TRAP INDEX OF ABUNDANCE, X_1 AND OF THE NUMBER OF UNITS OF GILLNET FISHING EFFORT, X_2 , (DATA FROM ROUNSEFELL AND KELEZ, 1938), FROM 1896 TO 1934, INCLUSIVE (NOTATION FOLLOWS SNEDECOR, 1946)

$n = 39 \quad \bar{y} = 2.5558 \quad \bar{x}_1 = 1.8920 \quad \bar{x}_2 = 334.9744 \quad \text{No. cycles} = 4$							
Source of variation	Sums of squares and products						
	D.F.	Sx_1^2	Sx_2^2	Sy^2	Sx_1x_2	Sx_1y	Sx_2y
Total	38	6,028,285	631,507	5.376020	1,269.551	4.771096	854.829
Cycles	3	1,577,233	26,227	1.097768	188.646	1.144999	152.229
Within cycles (error)	35	4,451,233	605,280	4.278252	1,080.905	3.626097	702.600
Correlation coefficients	$r_{y \cdot x_1} = 0.83097 \quad r_{y \cdot x_2} = 0.43664 \quad r_{x_1 \cdot x_2} = 0.65869 \quad R = 0.84387$						
Regression coefficients	$b'_{y \cdot x_1} = 0.95963 \quad b'_{y \cdot x_2} = -0.19536$						
Regression of Y	$\hat{Y} = 0.953100 + 0.940834X_1 - 0.000529X_2$						
Standard error of estimate	$S_{y \cdot 12} = 0.19319$						

Once $C'_{x_2}(o)$ had been obtained for any particular season, it could then be divided by k , 0.000529/.434, or 0.00122 to give the total number of fish reaching the river. However, although the estimates of the population so derived are correct in relative size to one another they do not yield the true population because the slope within years is not the same as the slope, b , between years, which is the only slope calculable from the available data.

A close approximation to the true populations can, however, be obtained by determining the relation between the calculated populations and estimates of the total population in one or more years. This was done by comparing the calculated estimates for the five years from 1941 through 1945 with the estimate of the total run to the river derived by adding the escapement estimated by the International Pacific Salmon Fisheries Commission to the gillnet catch on the river. The correlation,

TABLE 2
CALCULATED ESCAPEMENTS AND TOTAL RUNS IN THOUSANDS FOR THE FRASER
RIVER SOCKEYE SALMON¹

CYCLE A				CYCLE B			
Year	Escapement	Total run ²	Percent escapement	Year	Escapement	Total run	Percent escapement
1894	3431	7713	44.5	1895	3507	8658	40.5
1898	767	5837	13.1	1899	1431	12799	11.2
1902	1214	8393	14.5	1903	610	4863	12.5
1906	1251	5348	23.4	1907	418	2140	19.5
1910	1157	5613	20.6	1911	658	2837	23.2
1914	689	6382	10.8	1915	347	2172	16.0
1918	135	946	14.3	1919	318	1567	20.3
1922	456	1550	29.4	1923	442	1299	34.0
1926	1268	2650	47.8	1927	804	2587	31.1
1930	944	5532	17.1	1931	502	1936	25.9
1934	972	5992	16.2	1935	709	2119	33.5
1938	1277	5026	25.4	1939	364	1457	25.0
1942	2397	10682	22.4	1943	180	773	23.3
CYCLE C				CYCLE D			
1896	1196	5494	21.8	1897	4629	19051	24.3
1900	374	4760	7.9	1901	2372	28132	8.4
1904	389	2728	14.3	1905	3476	24157	14.4
1908	676	3426	19.7	1909	1636	22562	7.3
1912	1094	4457	24.5	1913	7157	38500	18.6
1916	117	1403	8.3	1917	435	7318	5.9
1920	431	1641	26.3	1921	354	2040	17.4
1924	549	1763	31.1	1925	619	2448	25.3
1928	314	1256	25.0	1929	617	2676	23.1
1932	678	2265	29.9	1933	469	2919	16.1
1936	2030	4763	42.6	1937	593	2189	27.1
1940	725	2553	28.4	1941	1392	4631	30.1
1944	542	2050	26.4	1945	454	2048	22.2

¹Subsequent to 1934 the data on number of gillnet licenses published by the Dominion Fisheries Department have been made comparable to the gillnet units of effort given by Rounsefell and Keeler (1938) by multiplying the number of licenses by 1.575. These data were used in making the above calculations.

²Escapement plus gillnet catch plus catch made before run reached the river.

r , between the two estimates is 0.9933. From this comparison it was obvious that by multiplying the calculated populations by 4.2896 a close approximation of the true population could be obtained so this procedure was followed.

The calculated escapements and total runs are shown in Table 2. It will be noted that the escapement for the "big" year of 1909 appears very low, especially in comparison to 1913. This is probably due to the fact that a large part of the escapement in 1909 and prior "big" years, as pointed out by Rounsefell and Kelez (1938) was derived from a very late fall run that was not fished by the fishery and consequently would not appear in the calculations. Certainly, the rough estimates of escapement available for the earlier years of the fishery do not yield any serious estimates of actual numbers. Even the 4,000,000 fish reported to have passed into Quesnel Lake in 1909, on careful reading are only an estimate made by counting for a short period each day and then estimating according to the number of hours. There is no mention of what portion of the day was chosen for counting!

The proportion that the escapement has formed of the total run has shown some interesting changes (see Table 2, and Figure 2). It may be noted that the proportion fell to the all-time low of 6 percent in the exceptionally intense fishery of 1917 during the first World War. After the partial obstruction of the upstream spawning migrations of 1913 and 1914, the proportion rose due partially to stricter regulation and partially to lack of interest in the small runs.

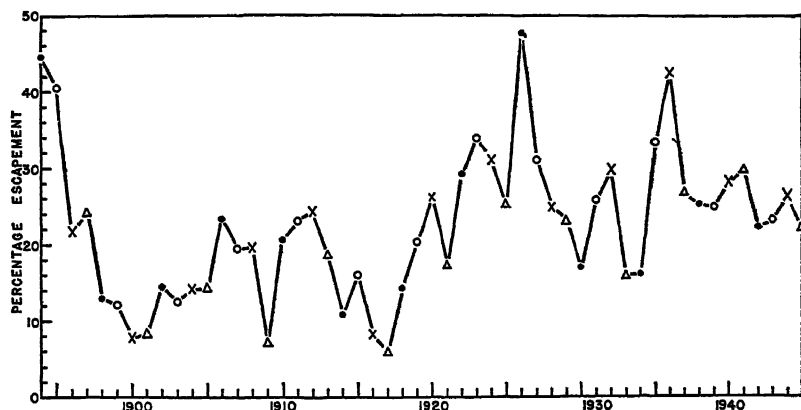


FIGURE 2

Showing the percentage that the escapement has formed of the total salmon run to the Fraser River, 1894 to 1945.

The runs of each of the 4-year age cycles are shown in Figure 3. The most striking feature is the decline of all but the "big" year cycle prior to the Hell's Gate slide of 1913-14, showing that overfishing played a major role in the decline of the Fraser River runs. Even in 1913, enough fish passed the Gate to bring back a large run, larger than all but a few off years, in 1917. Had fishing been curbed in that year the

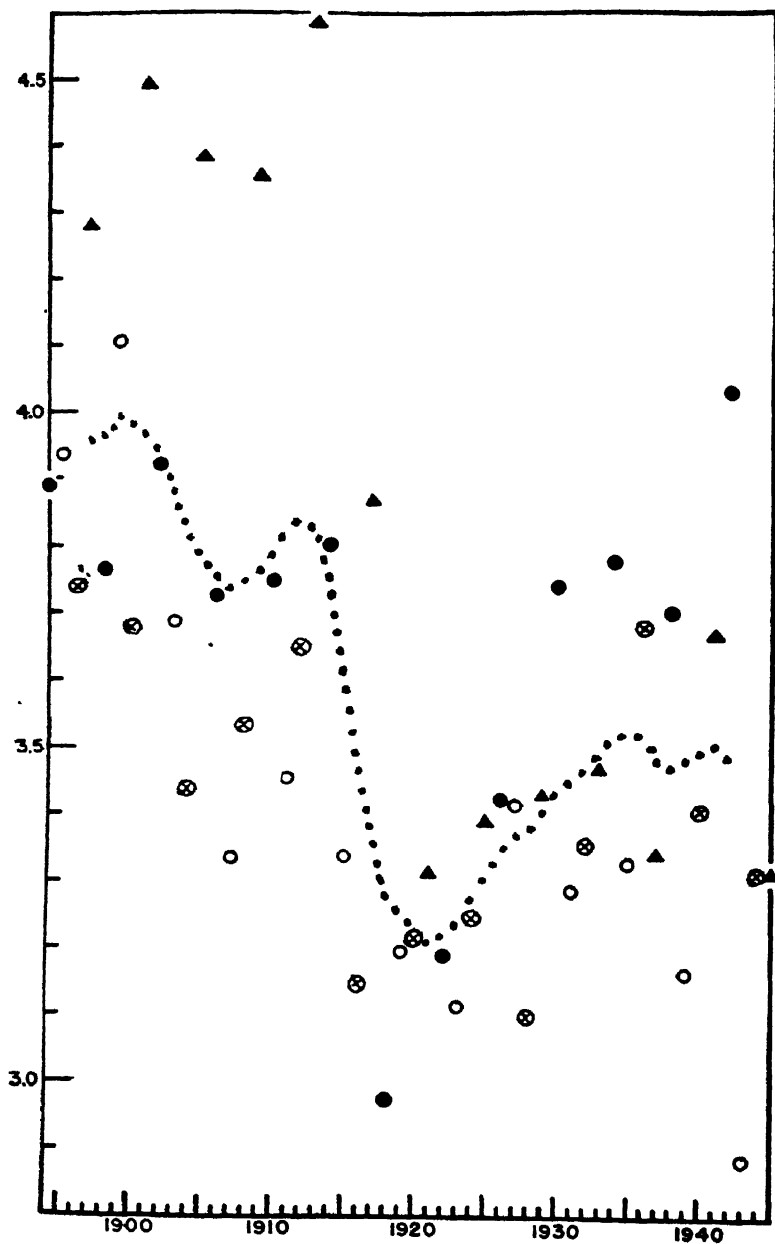


FIGURE 3

Showing the logarithms of the total runs by each age cycle and the trend of the geometric means smoothed by four. Data from Table 2. Cycle A equals black dots, Cycle B equals open circles, Cycle C equals crossed circles, and Cycle D equals black triangles.

"big" year cycle might have continued to dominate. Instead the fishery was so intense that the escapement was proportionately the lowest on record.

One major problem with two facets in the management of salmon runs is that of escapement. First, to what extent are variations in the size of the escapements associated with variations in the size of the runs? Second, what size of escapement will produce the largest surplus for the fishery out of the runs?

In order to answer the first question the calculated escapements and total runs for the Fraser River from 1894 to 1945 have been employed (Table 2). The total run and the escapement four years previous, are thus available for 48 pairs of observations.

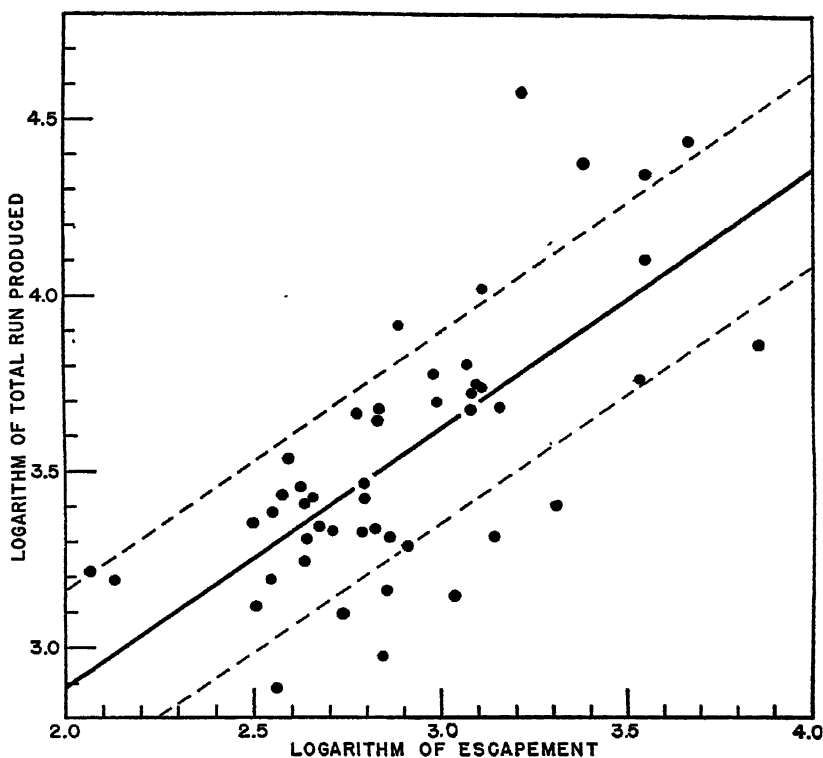


FIGURE 4

Showing the regression of the logarithm of the total run on the logarithm of the escapement four years previous. One standard error of estimate shown by dotted lines.

The size of the escapements and the size of the resulting runs show a high correlation, 0.7071. The coefficient of determination, 0.50, indicates that 50 per cent of the variation in the run is due to variation in the escapement. The pairs of observations and the regression line are plotted in Figure 4. It is obvious that for the Fraser River at least,

the size of the escapement is by far the dominant factor in determining the size of the runs.

However, there is a considerable range of variation in the returns from escapements of the same size which appears too large to be accounted for wholly by random variation. Foerster (1944) showed that predators can play an important role in the success of survival. Another likely source of variability is the great differences that exist in the distribution of spawners amongst the several major spawning areas in different years.

Thompson (1945) ascribed the major fluctuations in the success of reproduction in recent years to an obstruction in Hell's Gate Canyon at certain water levels of the upstream migration of adults bound for their spawning grounds.

This point can be tested by relating for the spawning years 1915 to 1941 the residuals of Y from the linear regression line shown in Figure 5 to the number of days given as passable at Hell's Gate in the report of Thompson. The correlation coefficient is 0.387 and a probability of 0.05 demands a coefficient of 0.381 with 27 degrees of freedom. The regression coefficient, $b_{y,x}$, is 0.00527, which when divided by its standard error 0.00251 yields a " t " of 2.1 which may have statistical significance.

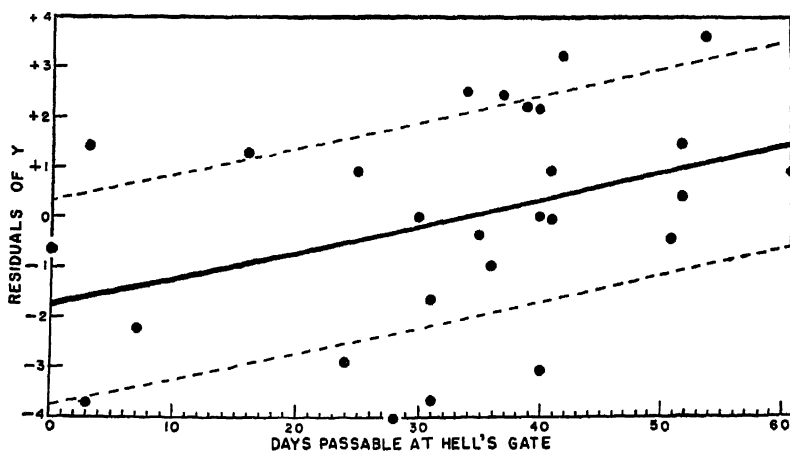


FIGURE 5

Showing the regression of the residuals of Y (See text and Figure 4) on the days passable at Hell's Gate

Perhaps it should be pointed out that even though the linear correlation coefficient of 0.387 were highly significant, it would mean only that 15 percent of the residual variability or only 7.5 percent of the total variability in the runs could be ascribed to the effects of water levels,

or of causes associated with water levels. Thus the data suggest a 7.5 percent effect of water levels at Hell's Gate on the success of spawning. However, since the runs above Hell's Gate have been smaller in the years with water level data available, and the effect is obscured by the inclusion in the data of the runs to the areas below Hell's Gate, it is possible that the effect is much larger than the data indicate.

The data on returns from escapements (Figure 4) indicate that the survival rate of the progeny decreases as the size of the escapement increases. Therefore, the largest difference between the size of the escapement and the number of returning salmon occurred when escapements were intermediate in size. This relationship has been repeatedly demonstrated in studies of population growth. However, it is not fully appreciated by the public in general. The clamor is for a return of the "good old days" when sockeye ascended the river in tremendous hordes. It is not generally realized that at very high population levels the efficiency of reproduction is so low that the major share of the run must spawn to maintain that level, leaving little surplus for the fishery.

Under the conditions prevailing in the Fraser River watershed during the 52 years from 1894 to 1945, the variation in returns is so great that any prediction is extremely hazardous. All that can safely be said is that the largest number of sockeye will be available for the fishery when the population is maintained at some optimum intermediate level of abundance.

The fact that there appears to be a maximum sustained harvest that could be taken under conditions prevailing during the past 52 years does not mean that there are no methods for increasing the harvest. Undoubtedly the lower efficiency of reproduction associated with large escapements has resulted partially from the overseeding of some spawning areas while others were underseeded. Regulation of the catch to permit larger proportionate escapements when the runs bound to underseeded watersheds are passing through the fishery might yield larger returns from the same number of spawners.

It should be borne in mind, however, that dominant cycles occurring every fourth year have been characteristic of the runs of sockeye to some of the lakes in the Fraser system. Further study will be necessary to determine whether it is desirable to seed these lakes every year. There is some possibility that such a procedure might produce lesser yields than the maintenance of dominant four-year cycles, (not all occurring on the same year) in different lakes.

Foerster (1944) made an excellent contribution in determining the effect of the control of predator and competing species in raising the survival rate of young sockeye in fresh water. As pointed out by

Rounsefell (1946) this method holds tremendous promise for the future.

A third method for increasing the returns per spawner consists in providing passage for salmon into lakes and streams now barren because of permanent stream obstructions. This method is being extensively employed at present by the Fish and Wildlife Service and the International Pacific Salmon Fisheries Commission.

In summary, a method has been shown for estimating the total run and the escapement in a salmon fishery, thus providing basic data necessary for intelligent management. It has been shown for one great river, the Fraser, that the size of each year's run is closely correlated with the number of spawners. It has also been shown that the total number of salmon that can be harvested from a run on a permanent basis cannot be increased beyond a certain point merely by increasing the number of spawners, unless either the spawners are so distributed over the watershed as to make better use of the available areas, or the environment of the nursery areas is changed.

REFERENCES

- Baranoff, Th. I.
 1918 On the question of the biological basis of fisheries Nauchnyi issledovatel'skii iktologicheskii Instit, *Izvestiia*, 1 (1), (1916): 81-128 (In Russian).
- DeLury, D. B.
 1947 On the estimation of biological populations. *Biometrics* 3 (4): 145-167.
- Dominion of Canada
 1947 Sixteenth annual report of the Department of Fisheries for the year 1945-46.
- Foerster, R. Earle
 1944 The relation of lake population density to size of young sockeye salmon (*Oncorhynchus nerka*). *Jour. Fish. Res. Bd., Canada*, 6: 267-280.
- Ricker, W. E.
 1940 Relation of "catch per unit effort" to abundance and rate of exploitation. *Jour. Fish. Res. Bd., Canada*, 5 (1): 43-70.
 1944 Further notes on fishing mortality and effort. *Copeia*, 1944.1, 23-44.
- Rounsefell, George A.
 1946 Fish production in lakes as a guide for estimating production in proposed reservoirs. *Copeia*, 1946.1, 29-40.
- Rounsefell, George A., and George B. Kelez
 1938 The salmon and salmon fisheries of Swiftsure Bank, Puget Sound, and the Fraser River. *Bull. U. S. Bur. Fish.*, pp. 693-823, Bull. No. 27, 1938.
- Snedecor, George W.
 1946 *Statistical Methods*, 4th Ed., Iowa State College Press, Ames, Iowa.
- Thompson, W. F.
 1945 Effect of the obstruction at Hell's Gate on the sockeye salmon of the Fraser River. *Inter. Pac. Salmon Fish. Comm.*, 175 pp.

THE ANALYSIS OF EXTINCTION TIME DATA IN BIOASSAY

K. MATHER

Department of Genetics, University of Birmingham

THE PROBLEM

THE STATISTICAL analysis of data from bioassays depending on quantal responses has received much attention during the last twenty years and the necessary analytical principles are now well understood. In general a transformation is sought of the proportions of tests on individuals subjected to treatment, which will yield a linear relation between proportion responding, as so transformed, and the treatment as measured on some suitable scale. The regression line defining this relation is then calculated and is used to find the treatment corresponding to some standard proportion of response chosen as convenient for purposes of comparison. Where the variation in treatment is a variation in dose administered, this standard is usually the dose giving the response in 50% of subjects. It is then designated as the Effective Dose 50 (*ED*50), or Lethal Dose 50 (*LD*50) where the response observed is death. Where time of exposure is the variable in treatment we have similarly Effective Time 50 (*ET*50) and Lethal Time 50 (*LT*50). The statistical analysis also provides means of testing the linearity of the regression relation, of comparing two or more regressions for correspondence, particularly in slope, and for finding the standard errors, and hence confidence limits, of the various quantities taken as specifying the regression line or measuring the potency of the treatment.

These methods have been developed particularly in relation to the type of assay where each treatment is administered to a different group of subjects, the fate of each of which can be observed individually. The experimental data then consist of proportions of individuals observed to respond to the treatment at its various levels and the analysis is undertaken by the now familiar method of probits, which has been fully described in both its derivation and its application by Finney (1947). The use of the probits in the comparison of bactericidal properties of a range of disinfectants has been illustrated by Berry and Michaels (1947, 1948, and in the press) who counted the numbers of bacteria surviving after exposure to these disinfectants for various lengths of time of samples from an original homogeneous culture.

Not all tests of bactericidal action however yield data of this kind. Another type of test, which we may refer to as the method of Extinction Times, aims at finding the time necessary to kill, or at least render ineffective, all the bacteria in the test sample. In practice a series of samples of the standard bacterial suspension, to which has been added the disinfectant, are "quenched" with nutritive broth after suitable intervals have elapsed. The samples are incubated, and any in which one or more active bacteria survive will then be detectable by the growth of the organism. The complete extinction of active bacteria can thus be related to the time of exposure to the disinfectant.

It has been supposed that there must exist a unique extinction time for any given bacterial suspension subjected to any given treatment of disinfection: that all samples quenched before the characteristic time had elapsed would show growth, and all after that time would be devoid of active bacteria. Such a simple view, of course, overlooks the variation which occurs between individual bacteria in their tolerance of disinfectants, and which must lead to marked sampling variation in the outcome of this type of test particularly. Instead of a simple sharp extinction point being shown, a series of tests may differ in the time of exposure after which no growth is found. And even in a single test growth may be found in samples which have been quenched after a longer exposure than others where no growth occurs. The series of 18 tests, in each of which samples were quenched every 2 minutes between exposure times of 12 and 26 minutes, shown in Table 1 illustrate the type of data which is yielded by the method of extinction times. Our problem is that of analysing such data so as to specify and measure in a way suitable for comparative purposes the relation between exposure time and extinction.

I am indebted to Professor H. Berry and Mr. H. S. Bean of the School of Pharmacy, University of London for drawing my attention to the problem, and to Mr. Bean also for providing me with the data of Table 1 as illustrative material.

THE METHOD OF ANALYSIS

The natural variation in individual tolerance of the bacteria will result in the death or inactivation of the organisms not being simultaneous. The number of survivors in a sample will not suddenly become zero, but will diminish more or less gradually as the period of exposure lengthens. This number will itself be subject to sampling variation, so that although there will be a mean number of survivors characteristic of a range of similar samples exposed for similar times, the number surviving will vary from sample to sample. Where, as in

TABLE 1
DISINFECTION OF *BACTERIUM COLI* (LISTER STRAIN 5933) BY 1.15% PHENOL

Series	Time of exposure in minutes							
	12	14	16	18	20	22	24	26
1	+	+	+	+	+	-	-	-
2	+	+	+	-	-	-	-	-
3	+	+	+	-	+	-	-	-
4	+	+	+	+	-	+	-	+
5	+	+	+	+	-	-	-	-
6	+	+	+	+	-	-	-	-
7	+	+	-	+	-	+	+	-
8	+	+	+	+	-	-	-	-
9	+	+	+	+	+	-	-	-
10	+	+	+	-	+	-	-	-
11	+	+	+	+	-	-	-	-
12	+	+	+	+	-	-	-	-
13	+	+	+	-	+	+	-	-
14	+	+	+	+	+	+	-	-
15	+	+	+	-	-	-	-	-
16	+	+	+	-	-	-	-	-
17	+	+	-	-	-	-	-	-
18	+	+	-	+	-	-	-	-
Total -ve	0	0	3	7	12	14	17	17
<i>p</i>	0.0	0.0	0.17	0.39	0.67	0.78	0.94	0.94
λ	-	-	1.79	0.94	0.41	0.25	0.06	0.06
<i>y</i>	-	-	0.58	-0.06	-0.90	-1.38	-2.86	-2.86

+ = positive reaction, i.e. bacteria surviving

- = negative reaction, i.e. no bacteria surviving

our tests, the survival of even one individual will lead to a positive result, the useful observations of response must cover ranges of samples in which the mean number of survivors is small. We shall in fact be concerned with the situation where comparable samples may contain only 0, 1, 2, 3 etc. individual survivors at the time of quenching. The frequencies of samples with these various numbers of survivors will then be expected to follow a Poisson distribution. Thus if the mean number of survivors is λ , the proportions of samples with 0, 1, 2, 3 etc. should fall in the series

$$e^{-\lambda} \left(1; \lambda; \frac{\lambda^2}{2!}; \frac{\lambda^3}{3!} \dots \right)$$

The technique distinguishes only between samples with on the one hand 0, and on the other 1 or more survivors, so that a proportion

$e^{-\lambda}$ should give the negative result of no growth and $1 - e^{-\lambda}$ the positive result of growth. The mean number of survivors is thus estimated by $-\log p$, where p is the proportion of negative samples.

Now the mean number may itself be taken as falling off logarithmically with time (or some simple function such as the logarithm of time) at least over the short range of assay times with which we are concerned. In other words $\log \lambda$ should be linearly related to time expressed on a suitable scale. Hence $\log \lambda = \log (-\log p)$ should give a straight line, within the limit of sampling error, when plotted against time. Instead, therefore of transforming into probits in order to achieve the desired linear relation, we must transform the data by taking the logarithm of the negative logarithm of the proportion of samples which fails to show growth after each of the exposure times used in the assay. This we will refer to as the loglog transformation for the sake of convenience. It may be remarked here that the transformation must be made using natural logarithms or the subsequent test of homogeneity of the samples and linearity of the relation with time will be vitiated.

Given this basic principle of transformation, the rest of the analysis with its weighting coefficients and working adjustments can be found by Fisher's method of maximum likelihood, applied in the way which Finney (1947) develops in his Appendix II. Where x is the time of exposure and $Y = \log (-\log P)$, P being the chance of a sample being *-ve*, we wish to find the constants α and β in the rectilinear relation

$$Y = \alpha + \beta x$$

The weight to be given in the calculation to any Y value can be found simply as the amount of information which the observed classification into *-ve* and *+ve* samples yields about Y . This is found as

$$I_1 = nS \left[\frac{1}{m} \left(\frac{dm}{dY} \right)^2 \right]$$

where m is the proportion in the class in question, n is the number of samples exposed for the given time and S indicates summation over all classes. Now $Y = \log (-\log P)$ so that $dP/dY = P \log P$, and $dQ/dY = [d(1 - P)]/dY = -P \log P$.

Then

$$r = n \left[\frac{1}{P} \left(\frac{dP}{dY} \right)^2 + \frac{1}{Q} \left(\frac{dQ}{dY} \right)^2 \right]$$

$$= n(P \log P)^2 \left(\frac{1}{P} + \frac{1}{Q} \right) = \frac{nP \overline{\log P^2}}{Q}$$

Where after a given exposure the chance of a sample being *-ve* is P , the probability of r samples being *-ve* out of n observed is

$$\frac{n!}{r!(n-r)!} P^r Q^{n-r}$$

With a series of exposure times the log likelihood of the particular set of results observed is proportional to

$$L = S[r \log P] + S[(n-r) \log Q]$$

summation proceeding over all times. Thus if α and β are the parameters determining Y and hence P their estimates will be given by the solution of the equations

$$\frac{\partial L}{\partial \alpha} = S \left[\frac{r}{P} \frac{\partial P}{\partial \alpha} \right] + S \left[\frac{n-r}{Q} \frac{\partial Q}{\partial \alpha} \right] = S \left[\frac{n(p-P)}{PQ} \frac{\partial P}{\partial \alpha} \right] = 0$$

$$\frac{\partial L}{\partial \beta} = S \left[\frac{r}{P} \frac{\partial P}{\partial \beta} \right] + S \left[\frac{n-r}{Q} \frac{\partial Q}{\partial \beta} \right] = S \left[\frac{n(p-P)}{PQ} \frac{\partial P}{\partial \beta} \right] = 0$$

where p is the observed proportion of *-ve* samples.

The direct estimation of α and β from this equation may well, as Finney points out, be impossible. They may, however, be obtained by a process of successive approximation. If, therefore, we obtain a first approximate relation between Y and time by inspection, adjustments to α and β can then be found from the general equations

$$\frac{\partial L}{\partial \alpha_1} + \delta \alpha \frac{\partial^2 L}{\partial \alpha_1^2} + \delta \beta \frac{\partial^2 L}{\partial \alpha_1 \partial \beta_1} = 0$$

$$\frac{\partial L}{\partial \beta_1} + \delta \alpha \frac{\partial^2 L}{\partial \beta_1 \partial \alpha_1} + \delta \beta \frac{\partial^2 L}{\partial \beta_1^2} = 0$$

the suffix to α and β denoting the substitution of the approximate values after differentiation.

Now

$$\frac{\partial Y}{\partial \alpha} = 1 \quad \text{and} \quad \frac{\partial Y}{\partial \beta} = x$$

so that

$$\frac{\partial P}{\partial \alpha} = P \log P \quad \text{and} \quad \frac{\partial P}{\partial \beta} = x P \log P$$

and the adjustment equation become

$$\delta \alpha S \left[\frac{nP \overline{\log P^2}}{Q} \right] + \delta \beta S \left[\frac{nP \overline{\log P^2}}{Q} x \right] = S \left[\frac{n \log P}{Q} (p - P) \right]$$

$$\delta \alpha S \left[\frac{nP \overline{\log P^2}}{Q} x \right] + \delta \beta S \left[\frac{nP \overline{\log P^2}}{Q} x^2 \right] = S \left[\frac{n \log P}{Q} (p - P)x \right]$$

Then substituting the weighting coefficient

$$w = \frac{I_Y}{n} = \frac{P \overline{\log P^2}}{Q}$$

$$\delta \alpha S(nw) + \delta \beta S(nwx) = S \left[nw \left(\frac{p - P}{P \log P} \right) \right] \quad (1)$$

$$\delta \alpha S(nwx) + \delta \beta S(nwx^2) = S \left[nw \left(\frac{p - P}{P \log P} \right) x \right] \quad (2)$$

Let

$$Y = \alpha + \beta x = a + b(x - \bar{x})$$

so that

$$a = \alpha + \beta \bar{x} \quad \text{and} \quad b = \beta$$

where

$$\bar{x} = \frac{S(nwx)}{S(nw)} \text{ is the weighted mean of } x$$

Equation (1) then becomes

$$(\delta \alpha + \delta \beta \bar{x}) S(nw) = \delta a S(nw) = S \left[nw \left(\frac{p - P}{P \log P} \right) \right]$$

Multiplying (1) by \bar{x} and subtracting the product from (2) gives

$$\begin{aligned} \delta \alpha S(nwx) - \bar{x} \delta \alpha S(nw) + \delta \beta S(nwx^2) - \bar{x}^2 \delta \beta S(nw) \\ = S \left[nw \left(\frac{p - P}{P \log P} \right) x \right] - \bar{x} S \left[nw \left(\frac{p - P}{P \log P} \right) \right] \end{aligned}$$

Now

$$\delta \alpha S(nwx) = \bar{x} \delta \alpha S(nw)$$

and

$$\bar{x}^2 \delta\beta S(nw) = \delta\beta \frac{S^2(nwx)}{S(nw)}$$

so that the last equation becomes

$$\delta\beta \left[S(nwx^2) - \frac{S^2(nwx)}{S(nw)} \right] = S \left[nw \left(\frac{p-P}{P \log P} \right) (x - \bar{x}) \right]$$

or

$$\delta\beta S[nw(x - \bar{x})^2] = S \left[nw(x - \bar{x}) \left(\frac{p-P}{P \log P} \right) \right]$$

Put

$$Y + \frac{p-P}{P \log P} = y_w,$$

the working loglog, and we find

$$a = \bar{Y} + \delta_a = \frac{S(nwY)}{S(nw)} + \frac{S \left[nw \left(\frac{p-P}{P \log P} \right) \right]}{S(nw)} = \frac{S(nwy_w)}{S(nw)} = \bar{y}_w$$

and

$$b = b_1 + \delta b = b_1 + \delta\beta = \frac{S[nw(x - \bar{x})Y]}{S[nw(x - \bar{x})^2]} + \frac{S \left[nw(x - \bar{x}) \left(\frac{p-P}{P \log P} \right) \right]}{S[nw(x - \bar{x})^2]} = \frac{S[nw(x - \bar{x})y_w]}{S[nw(x - \bar{x})^2]}$$

which are the required equations of estimation of the line relating response and time to the next approximation. The process can be repeated using the new calculated approximation to Y to obtain a third approximation, and so on as long as is necessary.

The analysis is thus exactly like a probit analysis except that we have

$$y = \frac{\log(-\log P)}{P \log P^2}$$

$$w = \frac{Q}{Q}$$

and

$$y_w = Y + \frac{p - P}{P \log P}$$

in place of the corresponding probit relations.

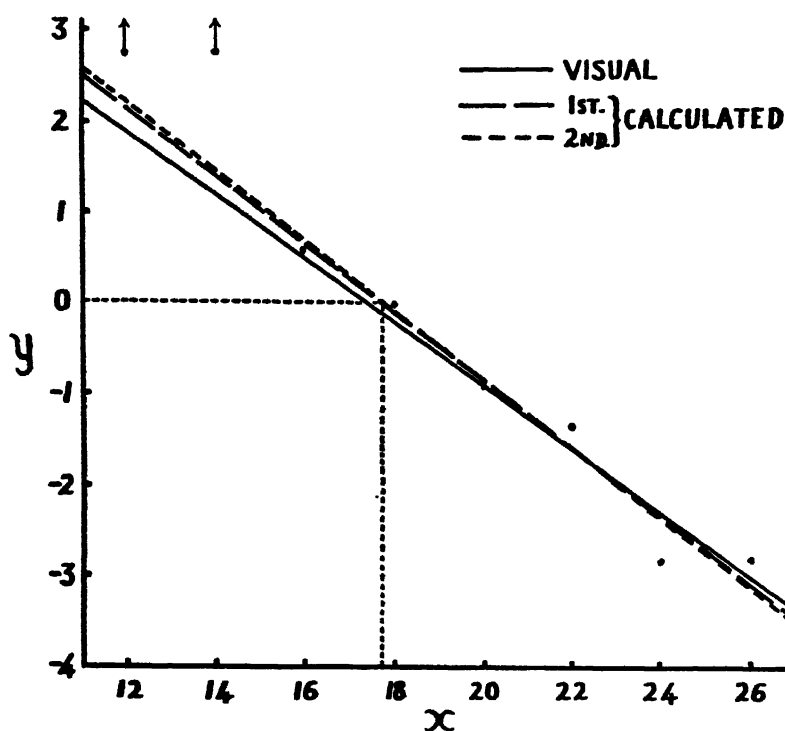


FIGURE 1

The calculation of the regression of $\log \log (y)$ on time (x) for the disinfection of *Bacterium coli* by 1.15% phenol. The observed results are shown as dots. The arrows to the dots at times 12 and 14 indicate that no negative cultures were obtained at these times so that the observed $\log \log$ is indeterminately large.

THE CALCULATION

The data of Table 1 will serve to illustrate the practical use of these equations of estimation. The data consist of 18 samples at each of the 8 times. The observed values of p , λ ($= -\log p$) and y ($= \log (-\log p)$) are shown at the bottom of the Table and y is plotted against the time x in Fig. 1. At times¹ 12 and 14 no $-ve$ samples were found so that y cannot be shown in the graph for these times. A trial line

¹Time has been expressed in minutes in this calculation. The scale of log minutes might be desirable on some occasions.

is drawn by inspection and for it we find the expected values, Y , as shown in Table 2. From these are found, by the counter transforma-

TABLE 2
CALCULATION OF THE LOGLOG REGRESSION LINE

x	p	y	Y	P	w	y_w	Y (2nd approx)
12	0 0000	—	1 9	0 00125	0 0539	2 0496	2 1631
14	0 0000	—	1 2	0 0361	0 4128	1 5012	1 4109
16	0 1667	0 5831	0 5	0 1923	0 6472	0.5807	0 6587
18	0 3989	-0 0571	-0 2	0 4410	0 5288	-0 0557	-0 0935
20	0 6667	-0 9038	-0 9	0 6661	0 3293	-0 9022	-0 8457
22	0 7778	-1 3812	-1 6	0 8172	0 1822	-1 3612	-1 5979
24	0 9444	-2 8612	-2 3	0 9046	0 0954	-2 7386	-2 3501
26	0 9444	-2 8612	-3 0	0 9514	0 0485	-2 8522	-3 1023

$$S(ux) = 40\ 4686$$

$$S(uy_w) = 0\ 135944$$

$$S(u) = 2\ 3001$$

$$S(ux^2) = 733\ 6140$$

$$S(uy_w^2) = 3\ 100673$$

$$S(uxy_w) = -5\ 73134$$

1st calculated regression line, from which Y (2nd approx) is found, is $Y = 6\ 6763 - 0\ 37610x$

tion $P = \text{antilog}(-\text{antilog } Y)$, the expected values of $\log P$, P and $Q = (1 - P)$. The weighting coefficients then follow as $(P \log P^2)/Q$. Knowing both the observed p and the first approximate expectation we can compute $(p - P)/(P \log P)$, which when added to Y gives y_w , the working loglog.

Table I, of the loglog transformation itself, and Table II of the weighting coefficients, maximum loglogs and working ranges, have been prepared to facilitate this part of the calculation. It will be observed that unlike probits, the relation of loglog to proportion is not symmetrical round 50% so that all these Tables must be constructed to cover the full range from $P = 0$ to 1.

Having obtained the values of y_w and w for each time x , the calculation proceeds as in probit analysis. One minor simplification will, however, often be possible and has been used in the present case. In all, 18 series of samples were run, each covering all 8 times. The weight of each of the 8 points in the calculation will thus be $18w$ where w is the appropriate coefficient. To save multiplying each of 8 w 's by 18, the actual w 's themselves have been used in the calculation. Since, where n is constant for all points,

$$a = \bar{y}_w = \frac{S(nwy_w)}{S(nw)} = \frac{S(wy_w)}{S(w)}$$

$$\bar{x} = \frac{S(nwx)}{S(nw)} = \frac{S(wx)}{S(w)}$$

Table I
TRANSFORMATION OF EXTINCTION POINT DATA TO LOGLOGS

$$Y = \log(-\log p)$$

p	0 000	0.001	0 002	0 003	0 004	0.005	0 006	0.007	0 008	0 009
0.00	∞	1.933	1.827	1.739	1.709	1.667	1.633	1.602	1.574	1.550
0.01	1.527	1.506	1.487	1.466	1.451	1.435	1.420	1.405	1.392	1.377
0.02	1.364	1.351	1.340	1.328	1.316	1.305	1.295	1.284	1.274	1.264
0.03	1.235	1.245	1.236	1.227	1.218	1.209	1.201	1.193	1.185	1.177
0.04	1.169	1.161	1.154	1.146	1.139	1.132	1.124	1.118	1.111	1.104
0.05	1.067	1.061	1.054	1.047	1.041	1.035	1.029	1.023	1.017	1.011
0.06	1.034	1.029	1.023	1.017	1.011	1.005	1.000	0.994	0.989	0.983
0.07	0.978	0.973	0.967	0.962	0.957	0.952	0.947	0.942	0.937	0.932
0.08	0.927	0.922	0.917	0.912	0.907	0.902	0.897	0.893	0.888	0.883
0.09	0.879	0.874	0.870	0.865	0.861	0.856	0.852	0.847	0.843	0.838
0.10	0.834	0.830	0.825	0.821	0.817	0.813	0.808	0.804	0.800	0.796
0.11	0.792	0.788	0.784	0.780	0.775	0.771	0.767	0.763	0.759	0.756
0.12	0.752	0.748	0.744	0.740	0.736	0.732	0.728	0.724	0.721	0.717
0.13	0.713	0.709	0.706	0.702	0.698	0.694	0.691	0.687	0.683	0.680
0.14	0.676	0.672	0.668	0.665	0.662	0.658	0.655	0.651	0.647	0.644
0.15	0.640	0.637	0.633	0.630	0.626	0.623	0.619	0.616	0.613	0.609
0.16	0.606	0.602	0.599	0.596	0.592	0.589	0.585	0.582	0.579	0.575
0.17	0.572	0.569	0.565	0.562	0.559	0.556	0.552	0.549	0.546	0.543
0.18	0.539	0.536	0.533	0.529	0.526	0.523	0.520	0.517	0.513	0.510
0.19	0.507	0.504	0.501	0.498	0.494	0.492	0.489	0.485	0.482	0.479
0.20	0.476	0.473	0.470	0.467	0.464	0.461	0.457	0.454	0.451	0.448
0.21	0.445	0.442	0.439	0.436	0.433	0.430	0.427	0.424	0.421	0.418
0.22	0.415	0.412	0.409	0.406	0.403	0.400	0.397	0.394	0.391	0.388
0.23	0.385	0.382	0.379	0.376	0.373	0.370	0.367	0.365	0.362	0.358
0.24	0.356	0.353	0.350	0.347	0.344	0.341	0.338	0.335	0.332	0.329
0.25	0.326	0.324	0.321	0.318	0.315	0.313	0.310	0.307	0.304	0.301
0.26	0.298	0.295	0.292	0.290	0.287	0.284	0.281	0.278	0.275	0.272
0.27	0.269	0.267	0.264	0.261	0.258	0.255	0.252	0.250	0.247	0.244
0.28	0.241	0.238	0.236	0.233	0.230	0.227	0.225	0.222	0.219	0.216
0.29	0.213	0.210	0.208	0.205	0.202	0.200	0.196	0.194	0.191	0.188
0.30	0.180	0.183	0.180	0.177	0.173	0.172	0.169	0.166	0.164	0.160
0.31	0.158	0.155	0.153	0.150	0.147	0.144	0.141	0.139	0.136	0.133
0.32	0.130	0.127	0.125	0.122	0.120	0.117	0.114	0.111	0.109	0.106
0.33	0.103	0.100	0.098	0.095	0.092	0.090	0.087	0.084	0.081	0.079
0.34	0.076	0.073	0.070	0.067	0.065	0.062	0.059	0.056	0.054	0.051
0.35	0.048	0.046	0.043	0.040	0.038	0.035	0.032	0.030	0.027	0.024
0.36	0.021	0.019	0.016	0.013	0.011	0.008	0.005	0.002	-0.000	-0.003
0.37	-0.006	-0.008	-0.011	-0.014	-0.017	-0.019	-0.022	-0.025	-0.028	-0.030
0.38	-0.033	-0.036	-0.038	-0.041	-0.044	-0.047	-0.049	-0.052	-0.055	-0.057
0.39	-0.060	-0.063	-0.066	-0.068	-0.071	-0.074	-0.076	-0.079	-0.082	-0.085
0.40	-0.087	-0.090	-0.093	-0.096	-0.098	-0.101	-0.104	-0.106	-0.109	-0.112
0.41	-0.115	-0.117	-0.120	-0.123	-0.126	-0.128	-0.131	-0.134	-0.137	-0.139
0.42	-0.142	-0.145	-0.148	-0.150	-0.153	-0.156	-0.159	-0.161	-0.164	-0.167
0.43	-0.170	-0.172	-0.175	-0.178	-0.181	-0.183	-0.186	-0.189	-0.192	-0.194
0.44	-0.197	-0.200	-0.203	-0.206	-0.208	-0.211	-0.214	-0.217	-0.219	-0.222
0.45	-0.225	-0.228	-0.231	-0.233	-0.236	-0.239	-0.242	-0.245	-0.247	-0.250
0.46	-0.253	-0.256	-0.259	-0.261	-0.264	-0.267	-0.270	-0.273	-0.275	-0.278
0.47	-0.281	-0.284	-0.287	-0.289	-0.292	-0.295	-0.298	-0.301	-0.304	-0.306
0.48	-0.309	-0.312	-0.315	-0.318	-0.321	-0.324	-0.326	-0.329	-0.332	-0.335
0.49	-0.338	-0.341	-0.343	-0.346	-0.349	-0.352	-0.355	-0.358	-0.361	-0.364

TABLE I—Continued

p	0 000	0.001	0 002	0.003	0 004	0 005	0 006	0 007	0 008	0.009
0 50	-0 366	-0.369	-0 372	-0 375	-0 378	-0 381	-0 384	-0 387	-0 390	-0.393
0 51	-0.395	-0 398	-0 401	-0.404	-0 407	-0.410	-0 413	-0 416	-0 419	-0 422
0 52	-0 425	-0 428	-0 431	-0 434	-0 436	-0 439	-0 442	-0 445	-0 448	-0.451
0 53	-0 454	-0 457	-0 460	-0 463	-0 466	-0 469	-0 472	-0.475	-0 478	-0 481
0 54	-0 484	-0 487	-0.490	-0.493	-0 496	-0 499	-0 502	-0 505	-0 508	-0.511
0 55	-0.515	-0.518	-0 521	-0 524	-0 527	-0 530	-0 533	-0 536	-0 539	-0.542
0 56	-0 545	-0.548	-0 551	-0 554	-0 557	-0 561	-0.564	-0 567	-0 570	-0 573
0 57	-0 576	-0 579	-0 582	-0 585	-0 589	-0 592	-0 595	-0 598	-0 601	-0 604
0 58	-0.608	-0.611	-0 614	-0 617	-0 620	-0 623	-0 627	-0 630	-0 633	-0 636
0 59	-0 639	-0 643	-0 646	-0 649	-0 652	-0 655	-0 659	-0.662	-0 665	-0 668
0 60	-0 672	-0 675	-0 678	-0 682	-0 685	-0 688	-0.691	-0.695	-0 698	-0.701
0.61	-0 705	-0 708	-0 711	-0 715	-0 718	-0.721	-0.725	-0.728	-0 731	-0 735
0.62	-0.738	-0.742	-0 745	-0 748	-0 752	-0 755	-0.758	-0.762	-0.765	-0 769
0.63	-0.772	-0.775	-0.779	-0.782	-0.786	-0.789	-0.793	-0.796	-0 800	-0.803
0.64	-0.807	-0.810	-0.814	-0 817	-0.821	-0.824	-0 828	-0.831	-0.835	-0.839
0.65	-0 842	-0.846	-0.849	-0.853	-0.856	-0.860	-0.864	-0.867	-0 871	-0.875
0.66	-0.878	-0.882	-0 886	-0.889	-0.893	-0.896	-0.900	-0.904	-0.908	-0.911
0.67	-0.915	-0 919	-0 923	-0.926	-0 930	-0.934	-0.938	-0.941	-0 945	-0 949
0.68	-0.953	-0 957	-0 961	-0 964	-0 968	-0 972	-0.976	-0 980	-0 984	-0 988
0.69	-0 991	-0 995	-0 999	-1 003	-1 007	-1 011	-1 015	-1 019	-1 023	-1 027
0.70	-1 031	-1 035	-1.039	-1 043	-1 047	-1 051	-1 055	-1.059	-1 063	-1 067
0 71	-1 072	-1 076	-1 080	-1 084	-1 088	-1 092	-1 096	-1 101	-1.105	-1 109
0.72	-1 113	-1.117	-1.122	-1.126	-1 130	-1 134	-1 139	-1 143	-1 147	-1 152
0.73	-1 156	-1 160	-1 165	-1 169	-1 173	-1 178	-1 183	-1 187	-1 191	-1 196
0.74	-1.200	-1.205	-1 209	-1 214	-1.218	-1.223	-1.228	-1.232	-1 237	-1 241
0 75	-1.246	-1.250	-1 255	-1 260	-1 264	-1 269	-1.274	-1 279	-1.283	-1.288
0 76	-1.293	-1.298	-1 303	-1 308	-1.312	-1 317	-1 322	-1 327	-1 332	-1 337
0 77	-1 342	-1.347	-1 352	-1.357	-1 362	-1.367	-1 372	-1 377	-1 382	-1 387
0 78	-1.392	-1.398	-1 403	-1.408	-1.413	-1 418	-1.424	-1 429	-1.434	-1 440
0 79	-1.445	-1 450	-1 456	-1.461	-1.467	-1 472	-1 478	-1.483	-1.489	-1.494
0 80	-1.500	-1 506	-1.511	-1.517	-1 522	-1.528	-1.534	-1.540	-1.546	-1 551
0 81	-1 557	-1 563	-1.569	-1.575	-1.581	-1.587	-1 593	-1.599	-1 605	-1 611
0.82	-1 617	-1 624	-1 630	-1 636	-1 642	-1.648	-1.654	-1 661	-1 667	-1.674
0 83	-1 680	-1 687	-1 693	-1 700	-1.707	-1.713	-1 720	-1 727	-1 733	-1 740
0.84	-1 746	-1.753	-1.760	-1 767	-1 774	-1 781	-1.789	-1.795	-1 802	-1 810
0 85	-1.817	-1 824	-1 831	-1 839	-1.846	-1 853	-1 861	-1 869	-1 876	-1.884
0 86	-1.892	-1 899	-1.907	-1 915	-1 923	-1 931	-1.939	-1.947	-1.955	-1 963
0 87	-1.971	-1.980	-1 988	-1 997	-2 005	-2 014	-2 022	-2 030	-2.039	-2.048
0.88	-2.057	-2 066	-2 075	-2.084	-2 093	-2.102	-2.112	-2 121	-2 130	-2 140
0 89	-2.150	-2 159	-2.169	-2 179	-2.188	-2 199	-2.209	-2.219	-2 229	-2 240
0.90	-2.250	-2 260	-2.271	-2 283	-2.294	-2 305	-2 316	-2 327	-2.338	-2 350
0.91	-2.361	-2 373	-2.385	-2.397	-2 409	-2 421	-2 434	-2.445	-2.458	-2 471
0.92	-2 484	-2 497	-2.511	-2 524	-2 537	-2 551	-2 565	-2 580	-2 594	-2 608
0 93	-2.623	-2.638	-2.654	-2 668	-2.684	-2 700	-2 717	-2.732	-2 749	-2 766
0.94	-2.782	-2.800	-2.817	-2 835	-2 854	-2.872	-2 891	-2 910	-2.930	-2 949
0.95	-2 970	-2.990	-3.012	-3.032	-3.055	-3.077	-3.101	-3.124	-3 149	-3.172
0.96	-3.199	-3.224	-3.249	-3 278	-3.305	-3.335	-3.364	-3.393	-3.427	-3.458
0.97	-3 490	-3.527	-3 561	-3.597	-3.634	-3.677	-3.717	-3 759	-3.803	-3 854
0.98	-3 902	-3.953	-4.006	-4.063	-4 129	-4.193	-4 262	-4 335	-4 415	-4 501
0.99	-4.595	-4 699	-4 828	-4.962	-5.116	-5.298	-5.521	-5 809	-6 215	-6.908

TABLE II

	Maximum	Range	Weighting Coefficient		Maximum	Range	Weighting Coefficient
Y	$Y - \frac{1}{\log P}$	$\frac{1}{P \log P}$	$\frac{P}{1-P} (\log P)^2$	Y	$Y - \frac{1}{\log P}$	$\frac{1}{P \log P}$	$\frac{P}{1-P} (\log P)^2$
1.00	2.0496	-114.9425	.0583	-1.05	1.8080	-4.0552	.2923
1.80	1.9653	-68.9655	.0879	-1.10	1.9039	-4.1911	.2804
1.70	1.8827	-43.4783	.1264	-1.15	2.0086	-4.3340	.2690
1.60	1.8019	-28.4091	.1755	-1.20	2.1201	-4.4863	.2580
1.50	1.7231	-19.7628	.2294	-1.25	2.2401	-4.6490	.2473
1.40	1.6466	-14.2450	.2897	-1.30	2.3697	-4.8193	.2369
1.30	1.5725	-10.6838	.3524	-1.35	2.5080	-5.0000	.2269
1.20	1.5012	-8.3195	.4141	-1.40	2.6552	-5.1894	.2174
1.10	1.4329	-6.7114	.4710	-1.45	2.8126	-5.3908	.2080
1.00	1.3679	-5.5741	.5221	-1.50	2.9823	-5.6022	.1990
0.95	1.3368	-5.1282	.5433	-1.55	3.1603	-5.8241	.1908
0.90	1.3066	-4.7551	.5657	-1.60	3.3529	-6.0606	.1822
0.85	1.2774	-4.4340	.5899	-1.65	3.5550	-6.3091	.1739
0.80	1.2493	-4.1597	.5998	-1.70	3.7735	-6.5703	.1665
0.75	1.2223	-3.9231	.6135	-1.75	4.0037	-6.8446	.1592
0.70	1.1966	-3.7202	.6247	-1.80	4.2496	-7.1378	.1522
0.65	1.1720	-3.5436	.6340	-1.85	4.5113	-7.4460	.1450
0.60	1.1488	-3.3944	.6403	-1.90	4.7845	-7.7640	.1380
0.55	1.1270	-3.2648	.6448	-1.95	5.0774	-8.1037	.1327
0.50	1.1066	-3.1546	.6470	-2.00	5.3910	-8.4602	.1265
0.45	1.0876	-3.0600	.6474	-2.10	6.0633	-9.2251	.1154
0.40	1.0703	-2.9789	.6462	-2.20	6.8253	-10.0806	.1049
0.35	1.0547	-2.9129	.6427	-2.30	7.6701	-11.0254	.0954
0.30	1.0409	-2.8571	.6378	-2.40	8.6254	-12.0773	.0865
0.25	1.0288	-2.8129	.6313	-2.50	9.6803	-13.2275	.0787
0.20	1.0187	-2.7770	.6237	-2.60	10.8390	-14.4928	.0712
0.15	1.0107	-2.7503	.6149	-2.70	12.1810	-15.9236	.0646
0.10	1.0047	-2.7322	.6047	-2.80	13.6474	-17.4825	.0593
0.05	1.0012	-2.7218	.5937	-2.90	15.2818	-19.1939	.0542
0.00	1.0000	-2.7181	.5820	-3.00	17.0803	-21.0970	.0494
-0.05	1.0013	-2.7218	.5695	-3.10	19.1222	-23.2558	.0432
-0.10	1.0032	-2.7315	.5563	-3.20	21.3098	-25.5102	.0400
-0.15	1.0118	-2.7473	.5420	-3.30	23.8003	-28.0890	.0359
-0.20	1.0214	-2.7693	.5280	-3.40	26.5401	-30.9598	.0335
-0.25	1.0340	-2.7972	.5146	-3.50	29.6126	-34.1297	.0303
-0.30	1.0499	-2.8321	.4999	-3.60	33.0300	-37.5940	.0260
-0.35	1.0690	-2.8711	.4853	-3.70	36.7858	-41.4938	.0246
-0.40	1.0919	-2.9163	.4705	-3.80	40.8429	-45.8621	.0226
-0.45	1.1184	-2.9674	.4559	-3.90	45.2030	-50.5051	.0200
-0.50	1.1488	-3.0239	.4412	-4.00	50.0448	-55.3556	.0166
-0.55	1.1834	-3.0864	.4263	-4.10	55.1410	-61.3497	.0182
-0.60	1.2222	-3.1546	.4119	-4.20	62.4667	-67.5676	.0134
-0.65	1.2657	-3.2289	.3976	-4.30	69.2294	-74.8269	.0148
-0.70	1.3137	-3.3091	.3835	-4.40	76.9008	-82.6446	.0082
-0.75	1.3669	-3.3956	.3695	-4.50	85.5901	-90.9091	.0091

TABLE II—Continued

	Maximum	Range	Weighting Coefficient		Maximum	Range	Weighting Coefficient
Y	$Y - \frac{1}{\log P}$	$\frac{1}{P \log P}$	$\frac{P}{1 - P} (\log P)^2$	Y	$Y - \frac{1}{\log P}$	$\frac{1}{P \log P}$	$\frac{P}{1 - P} (\log P)^2$
-0 80	1 4257	-3 4880	3559	-4 60	04 4099	-100 0000	0100
-0 85	1 4497	-3 5868	3427	-4 70	105 1901	-111 1111	.0111
-0 90	1 5394	-3 6928	.3295	-4 80	117.1512	-123 4568	0123
-0 95	1 6360	-3 8066	3168	-4 90	130.2351	-136 9863	.0135
-1 00	1 7181	-3 9262	.3044	-5 00	144.2537	-149 2537	.0000

$$b = \frac{S[nw(x - \bar{x})y_w]}{S[nw(x - \bar{x})^2]} = \frac{S[wy_w(x - \bar{x})]}{S[w(x - \bar{x})^2]}$$

the factor of $n = 18$ need never appear in their calculation. The sums of squares and products obtained using only w must, however, be multiplied by $n = 18$ to give the true values.

The data of Table 2 gives us

$$S(w) = 2.3001, \quad S(wx) = 40.4686, \quad S(wy_w) = 0.135944$$

and hence

$$\bar{x} = 17.5943, \quad \bar{y} = 0.05910$$

$$\begin{aligned} S[w(y_w - \bar{y}_w)^2] &= S(wy_w)^2 - \frac{S^2(wy_w)}{S(w)} \\ &= 3.100673 - 0.008035 = 3.092638 \end{aligned}$$

$$\begin{aligned} S[wy_w(x - \bar{x})] &= S(wxy_w) - \frac{S(wx)S(wy_w)}{S(w)} \\ &= -5.73134 - 2.39184 = -8.12318 \end{aligned}$$

$$\begin{aligned} S[w(x - \bar{x})^2] &= S(wx^2) - \frac{S^2(wx)}{S(w)} \\ &= 733.6140 - 712.0158 = 21.5982 \end{aligned}$$

Then

$$b = -\frac{8.12318}{21.5982} = -0.37610$$

and the sum of squares of y_u accounted for by the regression line is

$$\frac{18(-8.12318)^2}{21.5982} = 54.9930.$$

The total sum of squares of y_u is, of course, $18 \times 3.09264 = 55.6675$

The analysis of variance, or rather of χ^2 since in such a weighted analysis the sums of squares are χ^2 's, then becomes

<i>Item</i>	χ^2	<i>N</i>	<i>P</i>
Regression	54.9930	1	v. small
Remainder	0.6745	6	>0.99
<hr/>			
Total	55.6675	7	
<hr/>			

The experiment led to 8 observed proportions, so giving 7 degrees of freedom of which 1 corresponds to the χ^2 for regression and 6 to the remainder. The latter item has a very high probability and so affords no ground for regarding the data as inhomogeneous or the linear regression as inadequate.

Where the remainder χ^2 is not significant, we find the sampling variances of a and b as

$$V_a = \frac{1}{S(nw)} = \frac{1}{41.4018} = 0.02415$$

and

$$V_b = \frac{1}{S[nw(x - \bar{x})^2]} = \frac{1}{388.7676} = 0.002572$$

thus

$$a = \bar{y}_u = 0.0591 \pm 0.1554 \quad \text{and} \quad b = -0.3761 \pm 0.0507$$

Our new approximation to the loglog time relation is

$$\begin{aligned} Y &= 0.0591 - 0.3761 (x - 17.5943) \\ &= 6.6763 - 0.3761 x \end{aligned}$$

from which the values of Y in the last column of Table 2 are computed.

This second approximation can in turn be used as the basis for calculating the third approximation. In the present case the fresh calculation gives

$$a = -0.0469 \quad \text{and} \quad b = -0.3858$$

both of which values lie within the ranges covered by the standard errors of the values found by the first calculation. The remainder, or heterogeneity, χ^2 is indeed raised a little by the second calculation to 0.8988, so that clearly the second approximation is as good as the third and the second calculation redundant. Evidently the trial line fitted by eye and used as the basis of the first calculation was sufficiently good for the first calculation to provide an adequate correction. The three lines, visual trial, first calculated and second calculated are shown in Fig. 1. Their agreement is obviously close.

Having found the line relating loglog to time we can calculate any point on it which one may choose to regard as a convenient measure of its position and hence the potency of the disinfecting action, just as we find the *ED*50 as a convenient characteristic of the probit regression line. We could in fact take the exposure time which gave a proportion of 50% *-ve* samples, but this proportion has no special significance in terms of the bacteria—it corresponds to a mean number of 0.69 surviving bacteria per sample. It seems more appropriate to base comparisons on the time at which an average of 1 bacterium survives in the sample. Then $\lambda = 1$, and $P = 0.36788$ and $Y = 0$.

In the present case $Y = 0$ when $x = 17.751 \pm 0.414$ minutes the standard error being found from the formula, also used in probit analysis

$$s_{x_1} = (V_{x_1})^{1/2} = \left(\frac{1}{b^2} \left\{ \frac{1}{S(nw)} + \frac{(x_1 - \bar{x})^2}{S[nw(x - \bar{x})^2]} \right\} \right)^{1/2}$$

where x_1 is the value of x whose standard error it is required to know. The second calculation gives the single mean survivor time as 17.804 minutes, again well within a standard error of that from the first calculation.

Two disinfectant treatments can be compared in their action on a standard bacterial suspension through their single mean survivor times, just as by comparison of *ED*50's in probit analysis, provided the slopes of their loglog regression lines may be regarded as alike. Again this test of similarity of slope is made in a way exactly comparable to the corresponding test in probit analysis.

One final comparison of loglog and probit analyses. In the latter the most informative observations, the ones with the greatest weighting

coefficient, are those where $P = 0.50$ giving a probit value of 5.0. This is not true of loglogs. The amount of information per unit observation, *i.e.* the weighting coefficient w , is plotted against P in Fig. 2. For the

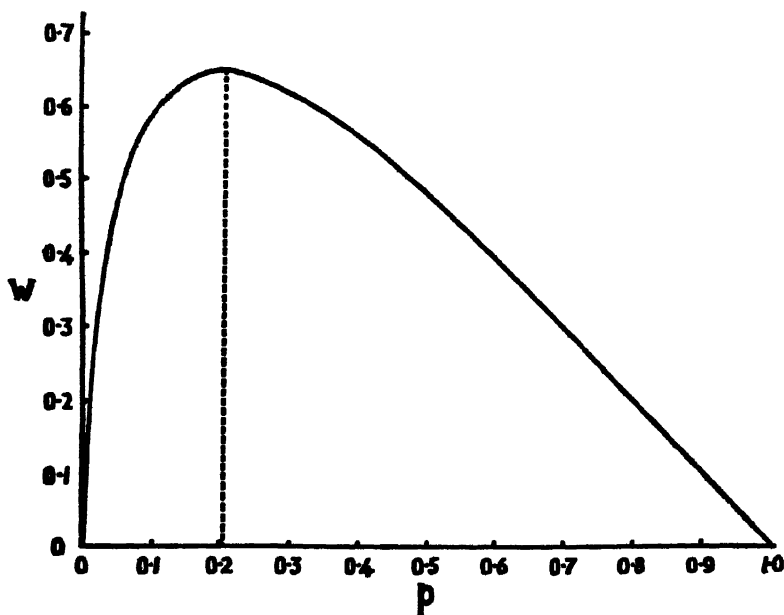


FIGURE 2

The relation of the amount of information or weighting factor (w) of the loglog to p the proportion of negative samples. $w = (p \overline{\log p})^2 / q$

most informative observation P lies between 0.2 and 0.3 with the maximum $w = 0.6476$ when $P = 0.2032$. This value of P is the same as the constant $e^{-\hat{m}}$ which Fisher (1947, Section 68) has found as the proportion of sterile samples yielding most information about the mean number of bacteria in the culture. He was especially concerned with estimation of the number of organisms by the dilution method, and he points out that a method depending on the mere recording of presence or absence of bacteria is at best much less informative than one which counts the bacteria in each sample. It would therefore appear that, observation for observation, the precision of the method of extinction times in bioassay is much less than can be achieved by plating and counting the survivors; though on the other hand it is of course also a much less time consuming method.

SUMMARY

In the method of extinction times as applied to disinfection of bacteria, the observational distinction is between samples which contain either none or 1 or more of active bacteria surviving after exposure to the disinfectant for given times. The data thus consist of proportions of samples failing to contain survivors after a series of exposure times. This proportion after any time is $e^{-\lambda}$ where λ is the mean number of survivors after that particular exposure. The mean is expected to fall off sufficiently nearly logarithmically with time. A rectilinear relation to time can thus be obtained by the loglog transformation $Y = \log(-\log P)$.

The calculation of the regression line relating loglog to time follows the same course as probit analysis, but with the weighting coefficient

$$w = \frac{P \overline{\log P}^2}{Q}$$

and the working loglog

$$y_w = Y + \frac{p - P}{P \overline{\log P}}$$

The loglog transformation also differs from the probit transformation in not being symmetrical round $P = 0.5$. For the most informative observations P has the value 0.2032, but any observation where P lies between 0.2 and 0.3 will closely approach this maximum in the information it yields.

A sample calculation is given and it is suggested that the place of the *ED*50 in probit analysis can be taken by the *single mean survivor time*, i.e. the time at which $\lambda = 1$, $Y = 0$ and $P = 0.3679$, in loglog analysis. The weighting coefficient is 90% of its maximum value at this proportion.

REFERENCES

- Berry, H. and Michaels, I. (1947, 1948) The evaluation of the bactericidal activity of ethyleneglycol and some of its monoalkyl ethers against *Bacterium coli*, Parts I-IV. *Quart. J. Pharm.* 20: 331-347, 348-366, 527-537; 21: 24-34.
Finney, D. J. (1947) *Probit Analysis*. Univ. Press Cambridge.
Fisher, R. A. (1947) *The Design of Experiments*. Oliver and Boyd, Edinburgh, 4th edn.

THE GENERAL THEORY OF PRIME-POWER LATTICE DESIGNS

III. THE ANALYSIS FOR p^3 VARIETIES IN BLOCKS OF p PLOTS WITH MORE THAN 3 REPLICATES.*

WALTER T. FEDERER†

INTRODUCTION

THE ANALYSIS FOR k^3 varieties or treatments in blocks of k plots for 3 replicates has been given by Yates [6]; k may be any of the integers 2, 3, 4, 5, 6, etc. In addition, he indicated the appropriate method of analysis for multiples of 3 replicates. At the same time he named this design a "three-dimensional lattice" while others [1, 2] have designated it as the "cubic lattice". In this paper the design will be known as a "3-dimensional lattice with one restriction", the restriction being that the whole block or replicate will be divided into p^3 (p = a prime number) incomplete blocks to which groups of p varieties are assigned at random.

The present paper is the third of a series of publications on prime-power lattice designs. The first two papers [4,5] presented the theory. The purpose of this paper is to illustrate, with a numerical example, the analysis for p^3 varieties in incomplete blocks of p varieties for more than 3 replicates. Although the numerical example contains 3^3 varieties in blocks of 3 plots with 4 replicates, the computational procedures are applicable for $p = 2, 3, 5, 7, 11$, etc. and for 4, 5, etc. replicates.

DESCRIPTION OF THE NUMERICAL EXAMPLE

Uniformity trial yield data [7] have been given for corn where the smallest unit of observation was pounds of ear corn for a 2×5 hill plot. The 4×5 and 2×10 hill plot yields were obtained by grouping 2 of the units of observations. The 2×10 hill plot yields were used to construct the numerical example (Table 1) illustrating the analysis for a 3^3 lattice in blocks of 3 plots with 4 replicates. The extension to p^3 varieties in blocks of p plots and to 5, 6, etc. replicates will be apparent from the explanation accompanying the present sample.

*Contribution of the Statistical Section of the Iowa Agricultural Experiment Station in co-operation with the Bureau of Agricultural Economics, U.S.D.A., Journal Paper No 1603, Project 890.

†Formerly Agricultural Statistician, Bureau of Agricultural Economics, collaborating with the Iowa Agricultural Experiment Station.

The incomplete block size was three 2×10 hill plots or 6×10 hills. The replicate size was 6×90 hills which may not be the most efficient shape of replicate for a randomized complete block design [3].

The varieties are designated as ijk where $i = 0, 1$, or 2 , $j = 0, 1$, or 2 and $k = 0, 1$, or 2 . The 27 variety numbers run from 000 to 222 (Table 1). The subscripts of the factors a , b , and c may be denoted as i , j , and k , respectively. This follows the notation of a true $p^3 = 3^3$ factorial experiment where the $3 = p$ levels of the 3 factors are in all possible combinations. The relationship is purely conventional and is useful in constructing an incomplete block design to determine which varietal comparisons are confounded with incomplete block differences. The factors a , b , and c are called pseudo-factors and the main effects and interactions pseudo-effects in an experiment which is not a true factorial but which makes use of factorial notation.

Groups of varieties (those making up a pseudo-effect) were assigned to the incomplete blocks at random and the variety or treatment designations were assigned to the 2×10 hill plots within the incomplete blocks at random. The field randomization and plot yields in pounds of ear corn per 2×10 hill plot are given in Table 1; the variety designations are given in parentheses. The incomplete block and replicate totals and the grand total are also given (Table 1). In the event that Table 1 is not constructed the above totals could be inserted in the field books in the appropriate places.

The pseudo-effects are obtained by taking certain combinations of varietal yields. Considering a single replicate, all $27 = p^3$ varietal yields are used to obtain the $3 = p$ levels of the pseudo-effect. The comparison among the $3 = p$ totals yields $2 = (p - 1)$ degrees of freedom for the particular pseudo-effect under consideration. For effect $(A)_0$, all the plots are summed for the pseudo-factor a , where $i = \text{zero}$; the varietal yields used to obtain $(A)_0$ are those listed under A for $i = \text{zero}$ in Table 2. For $(A)_1$ the varietal yields summed are those listed under A for $i = 1$ (Table 2) and for $(A)_2$ those listed under A for $i = 2$. The remaining main effects and interactions may be obtained in a similar manner (Table 2), if it is remembered that the powers and subscripts of the pseudo-effects are reduced to modulo $p = 3$ (that is, divided by $p = 3$ and the remainder substituted) [4].

In replicate I, the pseudo-effects, A , B , AB , and AB^2 are confounded with the differences among the incomplete blocks. Each pseudo-effect has $2 = (p - 1)$ degrees of freedom. There are $8 = (p^2 - 1)$ degrees of freedom confounded with the differences among the $9 = p^2$ incomplete blocks. The pseudo-effects A , C , AC , and AC^2 are confounded in replicate II; B , C , BC , and BC^2 in replicate III; and AB^2 , AC , BC , and

TABLE 1
FIELD ARRANGEMENT SHOWING PLOT YIELDS (POUNDS OF EAR CORN) FOR 3⁴ LATTICE IN BLOCKS OF 3 PLOTS WITH 4 REPLICATES
(VARIETY NUMBERS IN PARENTHESES)

Replicate I			Block Totals			Replicate II			Block Totals			Replicate III			Block Totals			Replicate IV			Block Totals			
(112)	(111)	(110)	92 0	(001)	(011)	(021)	100.5	(30 0)	(34 4)	(33 1)	(30 2)	(31 3)	(30 6)	(022)	(210)	(101)	92 1	(30 2)	(31 3)	(30 6)	(022)	(210)	(101)	92 0
30 6	32 0	30 3		33 0	34 4	33 1		(201)	(221)	(211)	(210)	(010)	(110)	(122)	(201)	(010)		(210)	(010)	(110)	(122)	(201)	(010)	
(002)	(000)	(001)	94 0	31 0	29 2	29 7	89.9	31 0	29 2	29 7	26 5	30 4	31 8	32 2	30 6	29 1	88 7	26 5	30 4	31 8	32 2	30 6	29 1	91 9
20 9	31 6	32 5		(120)	(110)	(100)		(120)	(110)	(100)	(002)	(102)	(202)	(020)	(102)	(211)		(002)	(102)	(202)	(020)	(102)	(211)	
(210)	(211)	(212)	92 6	29 0	30 7	30 2	90 8	29 0	30 7	30 2	21 2	20 4	29 3	28 1	28 7	30 5	82 0	21 2	20 4	29 3	28 1	28 7	30 5	87 3
32 5	30 6	29 5		(102)	(112)	(122)		(102)	(112)	(122)	(020)	(120)	(220)	(121)	(200)	(012)		(020)	(120)	(220)	(121)	(200)	(012)	
(120)	(122)	(121)	88 0	20 2	27 7	27 3	84 2	20 2	27 7	27 3	26 7	25 2	28 6	27 5	24 0	27 4	80 5	26 7	25 2	28 6	27 5	24 0	27 4	78 9
31 0	27 0	30 0		(202)	(212)	(222)		(202)	(212)	(222)	(201)	(101)	(001)	(112)	(221)	(000)		(201)	(101)	(001)	(112)	(221)	(000)	
(100)	(102)	(101)	96 1	27 4	28 8	26 8	83 0	27 4	28 8	26 8	25 9	26 8	24 4	23 7	22 5	23 0	77 1	25 9	26 8	24 4	23 7	22 5	23 0	69 2
32 6	30 6	32 9		(020)	(000)	(010)		(020)	(000)	(010)	(121)	(221)	(021)	(120)	(011)	(202)		(020)	(000)	(010)	(120)	(011)	(202)	
(011)	(010)	(012)	96 4	31 0	28 5	28 8	88 3	31 0	28 5	28 8	28 1	28 7	28 8	28 9	26 3	25 8	85 0	28 1	28 7	28 8	28 9	26 3	25 8	81 0
29 7	34 0	32 7		(012)	(002)	(022)		(012)	(002)	(022)	(212)	(112)	(012)	(001)	(110)	(222)		(212)	(112)	(012)	(001)	(110)	(222)	
(220)	(221)	(222)	98 2	31 5	33 9	30 6	96 0	31 5	33 9	30 6	31 5	29 9	29 5	26 3	28 6	27 0	90 9	31 5	29 9	29 5	26 3	28 6	27 0	81 9
31 1	34 0	33 1		(101)	(121)	(111)		(101)	(121)	(111)	(122)	(022)	(222)	(111)	(002)	(220)		(122)	(022)	(222)	(111)	(002)	(220)	
(020)	(021)	(022)	100 0	35 4	33 7	31 7	100 8	35 4	33 7	31 7	32 5	33 0	33 1	32 6	29 7	30 3	92 6	32 5	33 0	33 1	32 6	29 7	30 3	92 6
32 3	32 6	35 1		(200)	(210)	(220)		(200)	(210)	(220)	(211)	(011)	(111)	(100)	(021)	(212)		(211)	(011)	(111)	(100)	(021)	(212)	
(202)	(201)	(200)	96 4	33 1	29 3	31 2	93 6	33 1	29 3	31 2	31 6	29 3	30 8	29 9	31 8	28 1	91 7	31 6	29 3	30 8	29 9	31 8	28 1	89 8
33 8	31 4	31 2		Rep. total			827 1	788 1			761 6			3235 3										
33 8	31 4	31 2	855 5	Grand total			827 1	788 1			761 6			3235 3										

TABLE 2
MAIN EFFECTS AND INTERACTIONS (MOD 3) FOR THE PSEUDO-FACTORS OF THE TREATMENT COMBINATIONS $a_i b_j c_k$,
WHERE $i, j, k = 0, 1$, OR 2.

A $i =$			B $j =$			AB $i + j =$			AB^2 $i + 2j =$			C $k =$			AC $i + k =$			AC^2 $i + 2k =$		
0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
000	100	200	000	010	020	000	010	020	000	020	010	000	001	002	000	002	001	000	002	001
001	101	201	001	011	021	001	011	021	001	021	011	010	011	012	010	012	011	010	012	011
002	102	202	002	012	022	002	012	022	002	022	012	020	021	022	020	022	021	020	022	021
010	110	210	100	110	120	120	100	110	110	100	120	100	101	102	102	100	101	101	100	102
011	111	211	101	111	121	121	101	111	111	101	121	110	111	112	112	110	111	111	110	112
012	112	212	102	112	122	122	102	112	112	102	122	120	121	122	122	120	121	121	120	122
020	120	220	200	210	220	210	220	200	220	210	200	200	201	202	201	202	200	202	201	200
021	121	221	201	211	221	211	221	201	221	211	201	210	211	212	211	212	210	212	211	210
022	122	222	202	212	222	212	222	202	222	212	202	220	221	222	221	222	220	222	221	220

BC $j + k =$			BC^2 $j + 2k =$			ABC $i + j + k =$			ABC^2 $i + j + 2k =$			AB^2C $i + 2j + k =$			AB^2C^2 $i + 2j + 2k =$		
0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
000	001	002	000	002	001	000	001	002	000	002	001	000	001	002	000	002	001
100	101	102	100	102	101	012	010	011	011	010	012	011	012	010	012	011	010
200	201	202	200	202	201	021	022	020	022	021	020	022	020	021	021	020	022
010	010	011	011	010	012	102	100	101	101	100	102	102	100	101	101	100	102
112	110	111	111	110	112	111	112	110	112	111	110	110	111	112	110	112	111
210	210	211	211	210	212	120	121	122	120	122	121	121	122	120	122	121	120
021	022	020	022	021	020	201	202	200	202	201	200	201	202	200	202	201	200
121	122	120	122	121	120	210	211	212	210	212	211	212	210	211	211	210	212
221	222	220	222	221	220	222	220	221	221	220	222	220	221	222	220	222	221

TABLE 3
MAIN EFFECTS AND INTERACTIONS PER LEVEL (0, 1, 2) FOR EACH REPLICATE AND WEIGHTED (INVERSELY TO VARIANCE) MEAN EFFECTS AND INTERACTIONS

Treat- ment effect	Replicate I†			Sum	Replicate II†			Sum	Replicate III†			Sum	Replicate IV†			Sum	Weighted means per level (times $p^* = 9$)		
	0	1	2		0	1	2		0	1	2		0	1	2		0	1	2
(A) _n	290	4	277	9	287	2	284	5	250	0	204	7	248	7	253	0	262	9	248
(B) _n	286	5	281	0	287	1	281	7	259	1	271	8	264	7	248	8	250	2	250
(AB) _n	275	5	280	7	280	3	271	1	254	0	205	0	267	5	256	1	252	0	256
(A) _n	285	1	288	7	281	7	272	7	262	1	261	5	261	5	251	8	269	1	261
(C) _n	286	6	285	7	283	2	278	7	261	8	264	4	273	4	251	8	258	0	253
(A) _n	283	0	285	1	287	4	288	4	265	7	263	6	258	5	248	4	268	7	268
(A) _n	280	2	287	6	278	7	278	8	267	3	260	1	260	7	252	0	259	4	253
(B) _n	284	8	280	7	281	0	275	8	268	6	264	4	255	1	287	9	265	8	260
(BC) _n	283	8	287	7	284	0	272	3	268	4	267	2	258	6	250	8	253	0	254
(ABC) _n	287	5	280	3	277	7	271	0	259	8	265	5	262	8	249	9	251	4	250
(ABC) _n	291	2	281	0	283	3	272	4	250	3	262	0	265	9	248	8	274	8	248
(AB) _n	279	3	280	3	285	9	278	1	268	2	258	6	261	3	254	4	251	7	255
ABC _n	283	5	282	1	289	9	274	8	269	5	253	0	264	7	260	1	252	7	251

†Italics signifies that the effect was confounded with the block difference.

ABC^2 in replicate IV. A total of $32 = r(p^2 - 1)$ ($r = 4 =$ number of replicates) degrees of freedom are confounded in the 4 replicates. The confounding in replicates I, II, and III corresponds to that for Yates' [6] Z , Y , and X replicates, respectively.

Table 2 need not be constructed for each experiment but may be used for all succeeding experiments after it has once been constructed for a given value of p .

PROCEDURE FOR COMPUTATIONS

In addition to the totals obtained in Table 1, another table of totals is needed for the analysis of this and similar designs. The 3^2 yields in Table 1 which correspond to varieties listed under each level of a pseudo-effect in Table 2 were summed by replicates to obtain the totals in Table 3. Thus for the pseudo-effect $(ABC)_1$, the 9 varieties making up this total are (from Table 2)

001, 010, 022, 100, 112, 121, 202, 211, and 220.

The sum of the yields of these 9 varieties in replicate I (Table 1) is

$$32.5 + 34.0 + 35.1 + 32.6 + 30.6 + 30.0 + 33.8 + 30.6 + 31.1 = 290.3.$$

The remainder of the pseudo-effects were obtained in a similar manner. The method for obtaining the last 3 columns of Table 3 is explained in a later section of the paper. The main effects and interactions confounded in the various replicates are indicated by italics in Table 3.

The totals in Tables 1 and 3 and the unadjusted variety totals (Table 4) are all that are required to obtain the sums of squares for the analysis of variance.

I. CALCULATIONS FOR THE ANALYSIS OF VARIANCE

A procedure for obtaining the analysis of variance for a 3-dimensional lattice with one restriction is given below:

1. Correction term:

$$\frac{(\text{grand total})^2}{\text{total number}} = \frac{(3235.3)^2}{108} = 96,918.20 = CT.$$

2. Total sum of squares (from Table 1):

$$\begin{aligned} (30.6)^2 + (32.0)^2 + \dots + (31.8)^2 + (28.1)^2 - CT \\ = 97,680.03 - CT = 761.83 \end{aligned}$$

TABLE 4
UNADJUSTED TOTALS (POUNDS OF EAR CORN) FOR THE
27 VARIETIES IN TABLE 1

Variety number	Total (lbs.)	Variety number	Total (lbs.)
000	113.7	112	111.9
001	116.2	120	115.0
002	117.7	121	119.3
010	122.3	122	119.9
011	119.7	200	119.6
012	121.1	201	118.9
020	118.1	202	116.3
021	126.3	210	118.2
022	130.0	211	122.4
100	122.9	212	117.9
101	125.9	220	121.2
102	117.9	221	114.4
110	121.4	222	120.0
111	127.1		
		Total	3235.3

3. Replicate sum of squares (totals from Table 1):

$$\frac{(855.5)^2 + (827.1)^2 + (788.1)^2 + (764.6)^2}{27 = 3^3} - CT$$

$$= 97,099.61 - CT = 181.41.$$

4. Variety sum of squares (ignoring blocks)(totals from Table 4):

$$\frac{(113.7)^2 + (116.2)^2 + \dots + (120.0)^2}{4} - CT$$

$$= 97,033.42 - CT = 115.22.$$

5. The randomized block error sum of squares is obtained by subtraction of the replicate and variety sum of squares from the total,

$$761.83 - 181.41 - 115.22 = 465.20.$$

6. The sum of squares for blocks eliminating the varietal effect may be obtained as the sums of squares for the interaction of levels of the confounded pseudo-effects with replicates and the sum of squares for the com-

parisons of the mean confounded versus the mean unconfounded effects. In the present example the interblock error sum of squares will be derived from 3 sources which will be designated as components (a), (b), and (c).

(i) Component (a)

The component (a) sum of squares is the sum of the interaction sum of squares of the $3 = p$ levels 0, 1, and 2, of the effects A , B , AB^2 , C , AC , and BC with the replicates in which they are confounded. The interaction sum of squares for the AB^2 effect may be derived from the following 2-way table:

	Rep. I	Rep. IV
$(AB^2)_0$		
$(AB^2)_1$		
$(AB^2)_2$		

This interaction yields 2 degrees of freedom. There will be 2 from each of the 6 interactions or a total of 12 degrees of freedom for the component (a) sum of squares.

The within replicate sums of squares were obtained for all effects (Table 5); they may be used to obtain the interaction sum of squares. The within replicate I sum of squares for A is

$$\begin{aligned} \frac{1}{p^2} &= 9 \left((A)_0^2 + (A)_1^2 + (A)_2^2 - \frac{[(A)_0 + (A)_1 + (A)_2]^2}{3 = p} \right) \\ &= \frac{1}{9} \left((290.4)^2 + (277.9)^2 + (287.2)^2 - \frac{(855.5)^2}{3} \right) = 9.37. \end{aligned}$$

By making use of the totals in Table 3 the remainder of the within replicate sum of squares in Table 5 may be obtained. The within replicate and the replicate sums of squares should add to the total sum of squares within rounding errors.

The interaction sum of squares may be obtained from 2-way tables as described above or by adding the effect sum of squares within the replicates in which it is confounded and then subtracting the overall effect sum of squares. Thus for the interaction of $(A)_0$, $(A)_1$, and

TABLE 5
SUMS OF SQUARES FOR MAIN EFFECTS AND INTERACTIONS OF 3 LEVELS OF THE PSEUDO-FACTORS a , b , AND c WITHIN EACH OF THE 4 REPLICATES, AND FOR COMPONENTS (a) , (b) , AND (c)

MAIN EFFECT OR INTERACTION

Replicate No.	A	B	AB	AB ²	C	AC	AC ²	BC	BC ²	ABC	ABC ²	AB ² C ²	Total
I	9.37	1.80	15.68	2.72	0.09	1.08	7.11	4.23	1.07	9.73	6.86	6.81	380.47 (104 d.f., variety within replicate sum of squares)
II	18.61	6.00	0.44	1.50	45.06	43.88	2.80	1.08	2.01	11.03	2.22	4.34	
III	5.79	21.15	11.06	0.56	18.33	2.78	3.55	10.61	68.89	1.81	2.43	5.45	
IV	11.78	6.78	1.38	37.41	2.06	13.45	3.51	40.31	0.64	5.16	61.88	0.07	

COMPONENT (a) SUM OF SQUARES (Interaction of levels of a treatment with the reps. in which it is confounded).			COMPONENT (b) SUM OF SQUARES (Comparison by levels for mean unconfounded effect in 2 reps. and mean confounded effect in other 2 reps.)			COMPONENT (c) SUM OF SQUARES (Comparison by levels of mean unconfounded effect in 3 reps. and mean confounded effect in remaining rep.)		
s.s.	df		s.s.	df		s.s.	df	
A = 9.37 + 18.61 - 17.12 = 10.86	2		A	2	22.80	AB	2	6.30
B = 1.80 + 21.15 - 7.21 = 15.74	2		B	2	3.87	AB ²	2	5.54
AB ² = 2.72 + 37.41 - 26.92 = 13.21	2		AB ²	2	8.61	AC ²	2	5.45
C = 45.06 + 18.33 - 4.39 = 59.00	2		C	2	0.41	BC ²	2	56.70
AC = 43.88 + 13.45 - 32.06 = 4.67	2		AC	2	20.60	ABC ²	2	74.05
BC = 10.61 + 40.31 - 15.51 = 44.11	2		BC	2	6.54			
					71.42			122.62
		147.79		12				5

$(A)_2$ with replicates I and II (the effect A is confounded with the differences among incomplete blocks in replicates I and II) the sum of squares is

$$9.37 + 18.61$$

$$- \frac{(290.4 + 284.8)^2 + (277.9 + 275.8)^2 + (287.2 + 266.5)^2}{18 = 2p^2}$$

$$+ \frac{(855.5 + 827.1)^2}{54 = 2p^3} = 9.37 + 18.61 - 17.12 = 10.86.$$

The remaining interaction sums of squares (Table 5) for component (a) are obtained in a similar manner.

(ii) Component (b)

The component (b) sum of squares is the combined sums of squares for the comparison of the mean confounded level of the effect in 2 of the replicates and the mean unconfounded level of the effect in the other 2 replicates. For example, A is confounded in replicates I and II and unconfounded with the incomplete block differences in replicates III and IV. The sum of squares for this comparison is

$$\begin{aligned} & \frac{[(290.4 + 284.8 - 256.9 - 253.0)^2]}{p^2(1 + 1 + 1 + 1)} = 36 \\ & + \frac{(277.9 + 275.8 - 264.7 - 262.9)^2}{p^2(1 + 1 + 1 + 1)} = 36 \\ & + \frac{(287.2 + 266.5 - 266.5 - 248.7)^2}{p^2(1 + 1 + 1 + 1)} = 36 \\ & - \frac{(855.5 + 827.1 - 788.1 - 764.6)^2}{4p^3} = 22.30. \end{aligned}$$

The sum of squares for the effects B , AB^2 , C , AC , and BC are obtained similarly (Table 5). These comparisons yield a total of 12 degrees of freedom, 2 for each effect.

(iii) Component (c)

The component (c) sum of squares represents the sums of squares for the comparisons of the mean con-

founded effect in one replicate and the mean unfounded effect in the other 3 replicates. In the numerical example (Table 1) there are 4 effects, AB , AC^2 , BC^2 , and ABC^2 which are confounded in one replicate and unfounded in the other 3. Each comparison yields 2 degrees of freedom, making a total of 8 degrees of freedom for the component (c) sum of squares.

The sum of squares among the 3 differences adjusted for the mean difference for effect AB is

$$\begin{aligned} & \frac{[3(275.5) - 274.1 - 254.6 - 256.1]^2}{p^2(9 + 1 + 1 + 1)} = 108 \\ & + \frac{[3(290.7) - 276.7 - 266.0 - 252.0]^2}{p^2(9 + 1 + 1 + 1)} = 108 \\ & + \frac{[3(289.3) - 276.3 - 267.5 - 256.5]^2}{p^2(9 + 1 + 1 + 1)} = 108 \\ & - \frac{[3(855.5) - 827.1 - 788.1 - 764.6]^2}{12p^3} = 6.30. \end{aligned}$$

The sum of squares for the remaining 3 effects (Table 5) are obtained by a similar procedure.

7. The intrablock error sum of squares is obtained by subtracting the replicate, variety (ignoring blocks), and the block (eliminating varieties) sums of squares from the total sum of squares, thus

$$761.83 - 181.41 - 115.22 - 341.63 = 123.57.$$

The analysis of variance of the data in Table 1 is given in Table 6 for both the randomized complete block and the incomplete block designs.

An estimate of σ_e^2 is obtained as the intrablock error variance, 2.686 (Table 6). An estimate of the amount of intrablock information is given by

$$w = \frac{1}{\hat{\sigma}_e^2} = \frac{1}{2.686} = 0.3723008.$$

The total of the sum of squares for components (a), (b), and (c) gives the sum of squares for blocks (eliminating varieties) with 32 degrees of freedom (8 degrees of freedom are confounded with incomplete block

differences in each of the 4 replicates). The corresponding mean square has the expectation $\sigma_e^2 + (3/4) 3\sigma_b^2$ (see Kempthorne and Federer [4])

TABLE 6. ANALYSES OF VARIANCE
AS RANDOMIZED COMPLETE BLOCK DESIGN

Source of variation	d.f.	Sum of squares	Mean square	
Replicates	3	181.41	60.47	
Varieties	26	115.22	4.43	
Error	78	465.20	5.96	
Total	107	761.83		

AS INCOMPLETE BLOCK DESIGN WITH BLOCKS OF 3 VARIETIES

Source of variation	d.f.	Sum of squares	Mean square	Average value of Mean square
Replicates	3	181.41	60.47	
Blocks (elim. var.)	32	341.63	10.676	$\sigma_e^2 + (3/4) p\sigma_b^2$
Component (a)	12	147.59	12.299	$\sigma_e^2 + p\sigma_b^2$
Component (b)	12	71.42	5.952	$\sigma_e^2 + (2/4) p\sigma_b^2$
Component (c)	8	122.62	15.328	$\sigma_e^2 + (3/4) p\sigma_b^2$
Varieties	26	115.22	4.43	
Intrablock error	46	123.57	2.686	σ_e^2
Total	107	761.83		

where σ_e^2 is the expectation of the intrablock error variance and σ_b^2 is the expectation of the additional variance due to the variation among the incomplete block means freed of varietal effects. In general the expectation of blocks (eliminating varieties) mean square is $\sigma_e^2 + [(r-1)r]p\sigma_b^2$ where r is the number of replicates and p , a prime number, is the number of plots in the incomplete block. Using the blocks (eliminating varieties) mean square to obtain an estimate of the interblock error mean square which is equal to $1/w'$ or the reciprocal of the amount of interblock information, w' is estimated by

$$w' = \frac{1}{\hat{\sigma}_e^2 + 3\hat{\sigma}_b^2} = \frac{1}{\frac{4}{3}(10.676) - \frac{1}{3}(2.686)} = \frac{1}{13.339} = 0.0749681.$$

II. EFFICIENCY OF THE INCOMPLETE BLOCK DESIGN RELATIVE TO THE RANDOMIZED COMPLETE BLOCK DESIGN

The formula

$$\frac{2(p-1)}{p^3-1} \left\{ \frac{n_1}{w+(r-1)w'} + \frac{n_2}{2w+(r-2)w'} + \cdots + \frac{n_r}{rw} \right\},$$

where r = number of replicates, p^3 = number of varieties, n_1 = number of effects and interactions confounded in $(r-1)$ replicates,
 n_2 = number of effects and interactions confounded in $(r-2)$ replicates,
 \cdot
 \cdot
 \cdot

n_{r-1} = number of effects and interactions confounded in 1 replicate, and
 n_r = number of effects and interactions not confounded in any replicate,
 was given by Kempthorne and Federer [4] as the average error variance for the comparison of the mean difference of any two adjusted variety means. The average effective error variance per plot will be (no. of replicates = 4)/2 times this quantity. For the numerical example the average effective error variance is

$$\begin{aligned} & \frac{4}{13} \left\{ \frac{6}{2w+2w'} + \frac{4}{3w+w'} + \frac{3}{4w} \right\} \\ &= \frac{4}{13} \left\{ \frac{6}{0.8945378} + \frac{4}{1.1918705} + \frac{3}{1.4892032} \right\} \\ &= 3.716. \end{aligned}$$

The efficiency of this incomplete block design relative to the randomized complete design is the ratio of the two average variances, or

$$\frac{5.96}{3.716} = 160 \text{ percent.}$$

Since the efficiency of the incomplete block design is rather large (in this case the efficiency probably is inflated due to the choice of a long narrow replicate for the complete block) the adjustments of the variety means are expected to be appreciable and should be made. If the relative efficiency were small the adjustments for the variety means would be small and the adjusted variety means should differ little if any from the unadjusted means.

III ADJUSTED VARIETY MEANS

The mean yield of any treatment combination or variety yield, a, b, c_k , may be expressed [4] in terms of the means of the pseudo effects. The mean yield of the variety, a, b, c_k may be expressed as

$$\begin{aligned} & (A)_i + (B)_j + (AB)_{i,j} + (AB^2)_{i+2j} + (C)_k + (AC)_{i,k} + (AC^2)_{i-2k} \\ & + (BC)_{j,k} + (BC^2)_{j+2k} + (ABC)_{i+j+k} + (ABC^2)_{i+j-2k} \\ & + (AB^2C)_{i+2j+k} + (AB^2C^2)_{i+2j+2k} - \frac{(p^3 - p)\bar{x}}{(p - 1)} \end{aligned}$$

where \bar{x} is the mean of the experiment and the main effects and interactions are on a mean per plot basis. The subscripts i, j , and k or any combination give the level of the effect or interaction when all subscripts are reduced modulo p . The unadjusted or the adjusted means may be obtained from the above formula: the former may be obtained when the effect is given equal weights in all replicates and the latter when the effect is weighted inversely to the variance with which it is estimated in each replicate. To illustrate the use of the formula in obtaining the unadjusted variety totals (or means), the totals in Table 3 and the level of the effect in Table 7 are needed. The unadjusted total for variety 001 is

$$\begin{aligned} & 4[(A)_0 + (B)_0 + (AB)_0 + (AB^2)_0 + (C)_1 + (AC)_1 + (AC^2)_2 + (BC)_1 \\ & + (BC^2)_2 + (ABC)_1 + (ABC^2)_2 + (AB^2C)_1 + (AB^2C^2)_2 - 12\bar{x}] \\ & = \frac{1}{9} [1085.1 + 1069.1 + 1060.3 + 1063.6 + 1090.2 + 1075.7 \\ & + 1070.9 + 1092.8 + 1066.2 + 1082.5 + 1076.0 + 1074.2 \\ & + 1080.4] - 1437.9111 = 116.20 \end{aligned}$$

(or a mean of 29.050), where $1085.1 = 290.4 + 284.8 + 256.9 + 253.0$, etc.; and the divisor for each effect is 3^2 (p^2 in general) since there are 9 yields making up each level of the effect. The above total for variety 001 agrees with that obtained by adding the yields for this variety in each of the 4 replicates, which is the procedure usually followed in obtaining the unadjusted variety total yields.

Since some of the main effects and interactions are confounded with incomplete block differences in some of the replicates they will have a variance of $\sigma_i^2 + 3\sigma_b^2 = 1/w'$ in the replicates in which they are confounded. The unconfounded effects and interactions will be estimated with a variance $\sigma_i^2 = 1/w$. If the total of the level of an effect (Table 3) is weighted inversely to the variance with which it is estimated in the various replicates then a weighted total (of 9 plots) of the effects may be obtained (last 3 columns of Table 3). For example, A is confounded in replicates I and II and unconfounded in III and IV with estimates of variance in I and II of $\sigma_i^2 + 3\sigma_b^2 = 1/w'$ and of $\sigma_i^2 = 1/w$ in replicates III and IV. The weighted mean (of 9 plots) for $(A)_0$ is

$$\frac{w'(290.4 + 284.8) + w(256.9 + 253.0)}{2w + 2w'} = 260.4226,$$

where $w' = 0.0749681$ and $w = 0.3723008$. In a like manner the mean (of 9 plots) for $(ABC)_1$, which is unconfounded in all 4 replicates, is

$$\frac{w(290.3 + 272.3 + 265.5 + 254.4)}{4w} = 270.6250.$$

The remaining means of 9 plots (Table 3) are obtained in the same manner.

The adjusted variety means (or totals) are obtained from the formula given above when the weighted means (last 3 columns of Table 3) for the various levels of the effects are used. The particular level to use for each variety may be obtained from the first 14 columns of Table 7. These values need not be reproduced for each experiment but may be worked out for each value of p and then used again for all succeeding experiments. For variety 000, $i = 0$, $j = 0$, and $k = 0$ and the zero level for all effects is used to obtain the adjusted mean for this variety combination. For variety 001, $i = 0$, $j = 0$, and $k = 1$, the adjusted variety mean is obtained from the following levels of the effects,

$$\begin{aligned} & (A)_0 + (B)_0 + (AB)_0 + (AB^2)_0 + (C)_1 + (AC)_1 + (AC^2)_2 + (BC)_1 \\ & + (BC^2)_2 + (ABC)_1 + (ABC^2)_2 + (AB^2C)_1 + (AB^2C^2)_2 - 12\bar{x} \\ & = \frac{1}{9} [260.4226 + 265.9288 + 262.4743 + 266.8972 + 272.3838 \\ & + 272.5314 + 265.0869 + 278.5847 + 271.0529 + 270.6250 \\ & + 274.2139 + 268.5500 + 270.1000] - 359.4778 = 29.2835. \end{aligned}$$

The remainder of the adjusted variety means (Table 7) are obtained in the same manner.

If only the first 3 replicates had been used the design and analysis by the above computational procedure would result in the same adjusted variety means as the method described by Yates [6]. The computational procedure described above is more general than Yates' [6] method in some respects in that the method may be easily extended to the case of 4, 5 or more replicates for p^3 varieties.

IV. STANDARD ERRORS OF A DIFFERENCE BETWEEN ADJUSTED VARIETY MEANS

For the numerical example the average standard error of a mean difference between any 2 adjusted variety means is (see Kempthorne and Federer [4])

$$\sqrt{\frac{2(p-1)}{p^3-1} \left(\frac{n_1}{w+(r-1)w'} + \frac{n_2}{2w+(r-2)w'} + \cdots + \frac{n_r}{rw} \right)}$$

$$= \sqrt{\frac{2}{13} \left(\frac{6}{0.8945378} + \frac{4}{1.1918705} + \frac{3}{1.4892032} \right)} = 1.363,$$

where $r, n_1, n_2, \dots, n_r, w$ and w' are as defined previously.

Since this design is an unbalanced lattice design, some of the varieties occur together in an incomplete block while others do not. The standard error of a mean difference for any 2 varieties which occur together in an incomplete block is

$$\sqrt{\frac{2}{9} \left(\frac{3}{2w+2w'} + \frac{3}{3w+w'} + \frac{3}{4w} \right)} = 1.324.$$

The above standard error is applicable to such comparisons as the adjusted mean yield of variety 000 with any of the following adjusted variety means: 001, 002, 010, 020, 100, 200, 112, 221.

A second type of comparison, such as the adjusted mean yield of variety 000 with the adjusted means of any of the varieties, 012, 021, 102, 110, 201, 220, would have a standard error of a mean difference equal to

$$\sqrt{\frac{2}{9} \left(\frac{4}{2w+2w'} + \frac{4}{3w+w'} + \frac{1}{4w} \right)} = 1.374.$$

The third type of comparison for this design, which would involve such comparisons as the adjusted mean yield of variety 000 with the adjusted means of any of the varieties, 011, 022, 101, 111, 120, 121, 122, 202, 210, 211, 212, 222, would have a standard error of

$$\sqrt{\frac{2}{9} \left(\frac{5}{2w + 2w'} + \frac{2}{3w + w'} + \frac{2}{4w} \right)} = 1.383.$$

Since none of the above standard errors differ materially from the average standard error, 1.363 may be used as the standard error of a difference for the comparison of any 2 adjusted variety means.

V. COEFFICIENT OF VARIATION

The coefficient variation for the 3-dimensional lattice with one restriction is the square root of the average effective error mean square divided by the mean of the experiment. The coefficient of variation for the numerical example in Table 1 is

$$\frac{\sqrt{3.716}}{29.96} = 6.4 \text{ percent.}$$

LITERATURE CITED

1. Cochran, W. G. and Cox, G. M. Experimental Designs. Unpublished Manuscript, 1944.
2. Homeyer, P. G., Clem, M. A. and Federer, W. T. Punched Card and Calculating Machine Methods for Analyzing Lattice Experiments including Lattice Squares and the Cubic Lattice. *Iowa Agric. Expt. Sta. Res. Bul.* 347, 27-171, 1947.
3. Johnson, I. J. and Murphy, H. C. Lattice and Lattice Square Designs with Oat Uniformity Data and in Variety Trials. *Jour. Amer. Soc. Agron.* 35, 291-305, 1943.
4. Kempthorne, O. and Federer, W. T. The General Theory of Prime-power Lattice Designs I. Introduction and Designs for p^a Varieties in Blocks of p Plots. *Biometrics* 4, 54-79, 1948.
5. Kempthorne, O. and Federer, W. T. The General Theory of Prime-power Lattice Designs II. Designs for p^a Varieties in Blocks of p^a Plots, and in Squares. *Biometrics* 4, 109-121, 1948.
6. Yates, F. The Recovery of Inter-block Information in Variety Trials arranged in Three Dimensional Lattices. *Annals of Eugenics* 9, 136-156, 1939.
7. Zuber, M. A. Comparison of the Relative Efficiency of Various Experimental Designs for Corn Yield Tests. Unpublished M.S. thesis. Iowa State College Library, Ames, Iowa, 1940.

A RELATION BETWEEN THE LOGARITHMIC, POISSON, AND NEGATIVE BINOMIAL SERIES

M. H. QUENOUILLE

Marischal College, Aberdeen, Scotland

IN RECENT YEARS, many applications [3, 4] have been found for the logarithmic series developed by R. A. Fisher [1] in an investigation of the frequency distribution of numbers of species of animals obtained in random samples. Fisher derived this distribution by first considering a Poisson distribution with mean m , since this is the usually observed distribution where we are dealing with homogeneous material. However, where we are dealing with heterogeneous material, it is no longer possible to assume that m is fixed for all samples, so that Fisher assumed that it was distributed as a Gamma-type variable in the form

$$\frac{1}{(k-1)!} p^{-k} m^{k-1} e^{-m/p}.$$

With this assumption, we may consider the superposition of a set of Poisson distributions as resulting in one overall distribution: the negative-binomial distribution with parameter, $p/(1+p) = x$, say, and index k . The probability of observing a sample of size s is then

$$\frac{(k+s-1)!}{(k-1)!s!} \cdot \frac{p^s}{(1+p)^{k+s}}$$

or the coefficient of t^s in the expansion of

$$(1-x)^k(1-x)^{-k}$$

In particular, when $k \rightarrow 0$, this gave rise to a distribution whose first term tended to become infinite. However, upon excluding this term as being, in general, unobservable, Fisher obtained the logarithmic distribution:

$$\alpha x, \quad \frac{\alpha x^2}{2}, \quad \frac{\alpha x^3}{3}, \quad \dots$$

where $\alpha = -1/\log_e(1-x)$.

More recently [2], an alternative relation between these discrete

series was noted to exist and to be of some practical importance in bacterial counts where counts of individual bacteria and colony counts are taken. It was found that whereas the colony counts followed Poisson's distribution, the numbers of bacteria per colony were logarithmically distributed, and that, consequently, the bacterial counts were distributed in the negative binomial form. No proof of this relation was provided and it is not difficult to derive, but, since it is believed that its possible applications extend beyond the field of bacteriology, (e.g. quadrat sampling with over-dispersion), a simple proof is given below.

Suppose that the number of groups observed in any one occasion is distributed in the Poisson form, so that the probability of observing n groups is

$$P(n \text{ groups}) = \frac{e^{-m} m^n}{n!}$$

Then the probability of observing s individuals in any sample is

$$P(s \text{ individuals}) = \sum_{n=0}^{\infty} P(n \text{ groups}) \times P(s \text{ individuals in } n \text{ groups}).$$

Now the probability of observing s individuals in any one group is

$$\alpha x^s / s$$

or the coefficient of t^s in

$$-\alpha \log_e (1 - xt)$$

Likewise, the probability of observing s individuals in n groups is the coefficient of t^s in

$$[-\alpha \log_e (1 - xt)]^n$$

Thus, we have

$$\begin{aligned} P(s \text{ individuals}) &= \text{Coefficient of } t^s \text{ in } \sum_{n=0}^{\infty} \frac{e^{-m} m^n}{n!} \times [-\alpha \log_e (1 - xt)]^n \\ &= \text{Coefficient of } t^s \text{ in } \exp [-m - \alpha m \log_e (1 - xt)] \\ &= \text{Coefficient of } t^s \text{ in } (1 - xt)^{-\alpha m} e^{-m} \\ &= (1 - x)^{\alpha m} \frac{(\alpha m + s - 1)!}{(\alpha m - 1)! s!} x^s, \quad \text{since } (1 - x)^{-\alpha} = e \end{aligned}$$

This is the same as the $(s + 1)$ th term in a negative binomial series with parameter x and index αm . Consequently, the probability distribution of the number of individuals in random samples is the negative binomial.

Conversely, the assumption of any two of the distributions holding leads to the third distribution, provided that the parameters of the logarithmic and negative binomial distributions are equal when these are the known distributions.

Finally, it is worth noting that by this approach any disparity between the mean and the variance of a set of samples can be accounted for in terms of the parameter x . We have in fact,

$$\frac{\text{variance}}{\text{mean}} = \frac{1}{1 - x}.$$

This formula may be used to gauge x where over-dispersion is apparent.

REFERENCES

- [1] Fisher, R. A., Corbet, A. S., and Williams, C. B. The Relation between the Number of Species and the Number of Individuals in a Random Sample of Animal Population. *Journal of Animal Ecology* 12, 42-58, 1943.
- [2] Jones, P. C. T., and Mollison, J. E. A Technique for the Quantitative Estimation of Soil Micro-Organisms. *Journal of General Microbiology* 2, 54-69, 1948.
- [3] Williams, C. B. Some Applications of the Logarithmic Series and the Index of Diversity to Ecological Problems. *Journal of Ecology* 32, 1-44, 1944.
- [4] Williams, C. B. The Logarithmic Series and its Application to Biological Problems. *Journal of Ecology* 34, 253-271, 1947.

THE STATISTICAL ANALYSIS OF INSECT COUNTS BASED ON THE NEGATIVE BINOMIAL DISTRIBUTION

F. J. ANSCOMBE

Lecturer in Mathematics, Cambridge University, England

THIS NOTE GIVES a summary of the results of a mathematical investigation into the sampling theory of the negative binomial distribution, carried out during 1947 in the Statistical Department of Rothamsted Experimental Station. The work is a development of that of Fisher [5]. A full account will be given later elsewhere.

1. USE OF NEGATIVE BINOMIAL DISTRIBUTION

Insect counts in the field (and other population counts) are often fitted fairly well by a negative binomial distribution. This is described by two-constants, the mean m and the exponent k . The variance of the distribution is

$$(1) \quad m + \frac{m^2}{k},$$

the expected frequency of zeros is

$$(2) \quad p_0 = \left(1 + \frac{m}{k}\right)^{-k},$$

and the chance of observing any positive count r is

$$(3) \quad p_r = p_0 \binom{k+r-1}{r} \left(\frac{m}{m+k}\right)^r.$$

The Poisson distribution is obtained as the limit as $k \rightarrow \infty$. At the other end of the scale, as $k \rightarrow 0$, we approach the logarithmic series [6].

If we have several sets of counts on the same species of insect, from different districts or after different treatments, we may find that the mean m varies between the sets, but k remains approximately the same. To analyse such data statistically, we need to obtain a pooled estimate of k from all sets of counts and estimate the mean m separately for each set. There is some theoretical evidence [7] to show that k depends on the intrinsic power of the species to reproduce itself, while m depends on external factors. To try to fit negative binomial distributions with a common value of k to sets of counts on the same species is therefore a reasonable procedure.

2 ESTIMATION OF k FROM A SINGLE LARGE SAMPLE

We consider first the estimation of m and k from a single set of counts (made under uniform conditions). Suppose N counts have been made (i.e. the numbers of insects on N experimental units are counted), and n_0 of these counts are zeros (i.e. no insects were found on n_0 units). Let \bar{r} be the average number of insects found per count (i.e. the total number of insects counted, divided by N). Then \bar{r} is the best estimate of m . The best estimate of k , by the method of maximum likelihood or minimum χ^2 , is tedious to find; and in practice we require a shorter method. Three methods are useful and efficient in various circumstances.

- (i) We equate the variance of the sample to the variance of the distribution given above at (1). If s^2 is the sample variance, defined as the sum of squares of deviations of the N counts from \bar{r} , divided by $N - 1$, we get as our estimate of k

$$(4) \quad \frac{\bar{r}^2}{s^2 - \bar{r}}.$$

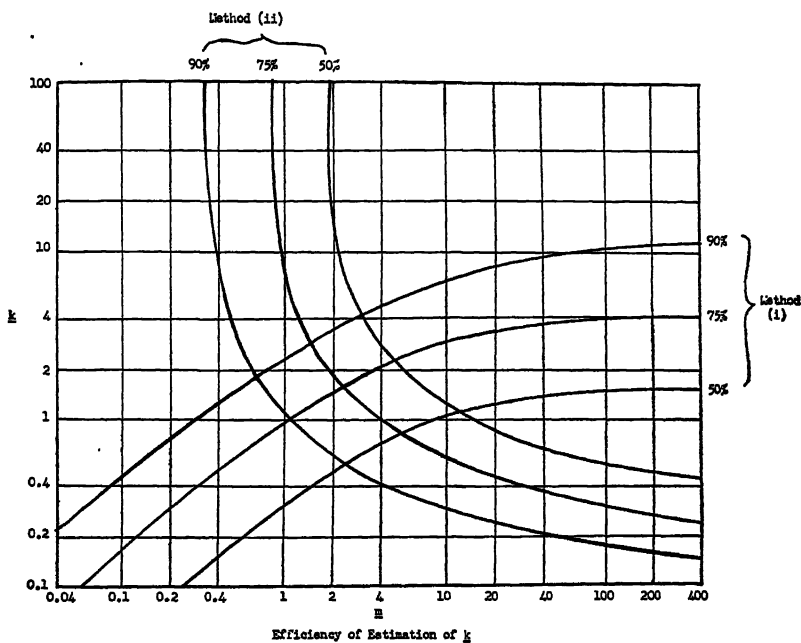
- (ii) We equate the observed proportion of zeros to the expected proportion given at (2) above, i.e. we choose k by successive approximation to satisfy

$$(5) \quad n_0 = N \left(1 + \frac{\bar{r}}{k} \right)^{-1}.$$

- (iii) We make a transformation of the actual counts r to a new variable y having a variance depending on k but not on m , and rather more nearly normally distributed [1]. We can then estimate k from the observed variance of y . The simplest transformation available is

$$(6) \quad y = \log_{10} (r + \frac{1}{2}k).$$

If k is between 2 and 5, this transformation may be used whenever m is not too small, say at least 15. If k is less than 2 or above 5, the transformation may still be used if m is large enough; but m may need to be considerably higher than 15 (so very much higher, in fact, when $k < 1$, that the possibility of use is almost ruled out). Under these conditions, the expected variance of y is approximately independent of m and equal to $0.1886 \psi'(k)$, where $\psi'(k)$ denotes the trigamma function, i.e. the second derivative of $\ln \Gamma(k)$ with respect to



k , and has been tabulated fully by Davis [4]. Roughly, $\psi'(k) = 1/(k - \frac{1}{2})$ when k is above 2, and $1/k^2$ if k is near 0. The procedure for finding k is to guess a value, use the above transformation (6), find the sample variance s_y^2 of y , equate this to the expected variance and so get a new estimate of k . The process is repeated if the new value of k is much different from the old one.

For k not less than 2, a more elaborate transformation may be used,

$$(7) \quad y = \text{Sinh}^{-1} \sqrt{\frac{r+c}{k-2c}}.$$

This has an expected variance of $0.25 \psi'(k)$. c is a constant; its best value is 0.375 if k is large, but somewhat smaller when k is small, 0.2 when $k = 2$. m may now be as low as 4 or 5. The transformation has not been investigated for $k < 2$.

Of these three methods, (i) and (ii) are quite easy to use, while (iii) is rather more bother. Roughly speaking, we use (i) if $k > 1$, (ii) if $k < 1$. The actual efficiency of the methods, as compared with the maximum

likelihood method, is indicated in the diagram, which shows, 90%, 75% and 50% efficiency contours for methods (i) and (ii). Method (iii) is only appropriate when m is not small, and then it is rather more efficient than (i).

The errors of estimation of m and k are independent, if N is large. \bar{r} is always a fully efficient estimate of m .

3 ESTIMATION OF k FROM SEVERAL SAMPLES

If we have several sets of counts and wish to estimate a common value of k , we may use developments of the three methods just described.

- (i) We guess a value of k , and calculate for each set of counts a quantity

$$(8) \quad T = \frac{(N-1)s^2 - (N-1-1/k)\bar{r}(1+\bar{r}/k)}{(\bar{r}+k)^2}.$$

Our object is to guess a value of k which makes the sum of these expressions T from all sets equal to zero. The process converges quite quickly if the working is suitably arranged. The divisor $(\bar{r}+k)^2$ is merely a weighting factor and can be replaced by \bar{r}^2 if \bar{r} is always rather larger than k (this making the working easier). It is assumed here that N is not very small. Presumably 10 would be large enough, but not 2.

- (ii) We guess a value of k and calculate for each set of counts a quantity

$$(9) \quad U = \log \left(1 + \frac{\bar{r}}{k}\right) \left[n_0 - \left(1 + \frac{\bar{r}}{k}\right)^{-k} \left\{ N - \frac{\bar{r}(k+1)}{2(\bar{r}+k)} \right\} \right].$$

Our object is to choose a value of k which makes the sum of these expressions U for all sets equal to zero. Again, the process converges quite quickly if the working is suitably arranged. The multiplier $\log (1 + \bar{r}/k)$ is a weighting factor, and may be replaced by $\log \bar{r}$ if \bar{r} is always large and much larger than k . In any case, it is not the optimum weighting factor which is much more troublesome to calculate. N is again assumed to be not very small.

- (iii) We calculate the variance of the transformed variable y for each set, pool the answers and equate to the theoretical variance. The method applies even if N is as small as possible, namely 2. (No estimate of k could be

derived from a single observation.) It is, however, subject to the restrictions mentioned above on the values of m and k for a suitable transformation to exist.

4. DESIGN AND ANALYSIS OF EXPERIMENTS

Let us consider an experiment in which t treatments are compared in "randomized blocks" of tN experimental units, these units having been divided at random into t sets of N units for the application of the treatments. There may be one or several such blocks. The observations consist of counting the number of insects on each experimental unit. If negative binomial distributions with a common value of k are to be fitted to the sets of N counts for each treatment in each block, k will be estimated by one of the methods just described. The analysis will then proceed on the totals of each set of N counts. The totals have negative binomial distributions with exponent equal to Nk , and may be transformed as already explained to permit of an analysis of variance. (The transformation is different from what may have been used in determining k , since now the exponent is Nk). The residual error mean square in the analysis of variance may be compared with the expected variance for the transformation, given above, as a test of heterogeneity in the observations.

If the estimation of k is by method (iii), N may be as small as 2, but the infestations must not be too low. For methods (i) and (ii) N will need to be larger, and in the absence of more precise information it seems reasonable to recommend that N should be at least 10. These remarks amplify Beall's suggestion [3] that N should be at least 2 in experiments of this kind, so that k can be estimated.

If it is desired to avoid the assumption of a negative binomial form of distribution, with constant exponent k , it would be possible to proceed by method (iii) to derive an estimate of k and then use the transformation so defined for all further work. This would probably be as satisfactory a transformation as could be used, unless some precise assumption (other than the negative binomial one) were made about the distribution of the observations. The final analysis of variance would then be based on the totals of the individual transformed counts per set of N units and not on a direct transformation of the total count from the N units. In fact, the use of the transformation

$$(10) \quad y = \log(r + 1)$$

in this way is well known and common where the standard deviation of r appears to be roughly proportional to the mean.

TABLE

Number of eggs on ten shoots	\bar{r}	U T (assuming $k = 0.5$)	
0^{10} (27 sites)	0.0	0.00	0.0
$0^8, 1$ (7 sites)	0.1	0.00	0.2
$0^8, 2$ (3 sites)	0.2	0.11	3.3
$0^8, 1^2$	"	-0.04	-0.7
$0^8, 3$ (2 sites)	0.3	0.27	7.4
$0^8, 1, 1$	0.5	0.36	7.5
$0^8, 2, 3$	"	0.36	3.5
$0^8, 1^3, 2$ (3 sites)	"	-0.24	-2.5
$0^8, 1^5$	"	-0.54	-4.5
$0^8, 6$ (2 sites)	0.6	0.87	19.1
$0^8, 1^2, 2^2$	"	-0.16	-2.3
$0^8, 1^3, 2^2$	0.7	-0.45	-3.9
$0^8, 1^2, 2, 1$	0.8	0.04	0.6
$0^7, 2^2, 5$	0.9	0.59	3.7
$0^8, 1, 2, 3, 1$	1.0	0.25	-0.4
$0^8, 5, 6$	1.1	1.36	9.5
$0^7, 1, 5, 9$	1.5	1.37	10.6
$0^4, 1^2, 2^2, 4, 5$	"	-0.43	-3.4
$0^8, 3, 13$	1.6	2.12	23.9
$0^8, 1, 2^2, 8, 9$	2.2	0.70	3.1
$0^8, 2, 4, 5^2, 6$	"	0.70	-3.5
$0^8, 1, 3, 6, 16$	2.6	1.77	12.6
$0^8, 1, 2, 4^2, 17$	2.8	1.11	10.9
$0^8, 29$	2.9	4.51	53.5
$0^8, 1^2, 2, 12, 17$	3.3	1.42	10.7
$0^8, 1^3, 2, 1, 6, 8, 10$	3.3	-1.23	-4.3
$0^8, 2, 10, 11, 12$	3.5	2.43	3.2
$0^8, 16, 24$	4.0	4.66	20.7

5. A NUMERICAL EXAMPLE

I have not encountered any experimental observations of the sort just considered. (Beall [3] gives some examples, however). To illustrate the methods of §3, I give here some counts of eggs of *Aphis fabae* made by Dr. D. Price Jones in the course of a survey of the Eastern Counties of England in 1947, which he has kindly allowed me to reproduce. Ninety-four hedgerow spindle sites were visited, that had been cut down the previous winter, so that the shoots were of one-year growth. At each site ten shoots were removed and the *A. fabae* eggs on them subsequently counted. The counts are shown in the table arranged in order of in-

TABLE (Continued)

Number of eggs on ten shoots		U (assuming k	T 0.5)
$0^4, 1, 2^2, 3, 7, 26$	4.1	0.89	14.7
$0^3, 3^2, 4, 5, 7, 9, 12$	4.3	-0.01	-6.1
$1, 2^2, 4^2, 5^2, 6, 8, 13$	5.0	-2.92	-9.1
$0^5, 1, 3, 5, 6, 40$	5.5	2.49	25.2
$0^2, 1^3, 2^2, 3, 8, 13, 33$	6.3	-0.59	7.4
$2^2, 3, 4, 5^2, 6, 9, 14^2$	6.4	-2.86	-9.2
$0^2, 1, 2, 3, 10, 11^2, 12, 17$	6.7	-0.52	-6.5
$0^2, 1, 3^2, 4, 6, 11, 15, 25$	6.8	-0.51	-2.2
$1, 2^3, 3, 6, 11, 14^2, 19$	7.4	-2.80	-7.0
$1^3, 2^2, 4, 6, 14, 47$	8.0	-2.77	12.2
$0^2, 1, 2, 4, 6^2, 9, 31^2$	9.0	-0.17	1.0
$1^3, 2, 3, 4, 10, 13, 31, 35$	10.1	-2.67	-0.3
$0^2, 3^2, 6, 7, 15, 19, 21, 35$	10.9	0.08	-4.4
$0, 3^2, 4, 5, 6, 13, 15, 25, 35$	"	-1.28	-4.5
$0^2, 1, 2, 4^2, 6, 23, 27, 50$	11.7	0.17	3.1
$0^4, 2, 3, 10, 12, 42, 50$	11.9	2.98	6.8
$1, 3^2, 4^2, 6, 10^2, 39, 45$	12.5	-2.57	0.0
$2^2, 3^2, 5, 11, 13, 18, 66$	"	-2.57	7.0
$0^2, 2, 6, 7, 10, 11, 12, 17, 83$	14.8	0.48	9.7
$0^2, 6, 10, 11, 18, 21, 35, 65$	16.6	2.17	-0.9
$2, 4^2, 11^2, 19, 20, 31, 32, 39$	17.3	-2.41	-8.7
$3, 10, 14, 15, 17^2, 23, 24^2, 33$	18.0	-2.39	-11.8
$0^2, 3^2, 5, 7, 13, 19, 148$	19.8	2.49	31.5
$4, 9, 12, 17, 18, 22, 23, 24, 34, 70$	23.3	-2.25	-8.3
$22, 24^2, 31, 34, 36, 43, 44, 48, 58$	36.4	-2.01	-12.9
$0, 1, 3, 17, 33, 38, 48, 49, 84, 110$	38.3	-0.10	-5.8
$1, 8, 10, 18, 26, 32, 44, 52, 82, 120$	39.3	-1.97	-5.9
$21, 35, 51, 59^2, 70, 105, 120, 123, 163$	80.6	-1.61	-11.1

creasing \bar{r} . The eighth line, for example, indicates that at three sites six shoots had no eggs, three had one and one had two eggs; while the following line indicates that at one site there were five shoots without eggs and five with one egg.

The values of U and T are given on the assumption that $k = 0.5$. The sum of the U s is nearly zero, and 0.5 is close to the estimate of k given by this method. For the range of values of m that appears to have been encountered, method (ii) is considerably more efficient than method (i), while method (iii) is inappropriate. We should therefore accept the value of k given by method (ii), if any.

It appears, however, when we plot U against \bar{r} , that the value of k is not constant but increases with m . Thus, if we consider the counts in which \bar{r} exceeds 4.0, method (ii) gives k equal to about 0.65; while for the counts in which \bar{r} is less than 4.0 k is in the neighbourhood of 0.3. The effect is too marked to be attributed to the negative correlation between n_0 and \bar{r} that occurs in repeated sampling of the same population. We observe a similar increase in k if we use method (i), plotting T against \bar{r} ; but now there is further cause of perplexity, in that the values of k indicated by method (i) are appreciably lower than those of method (ii). Method (i) indicates an overall value for k round about 0.35 and 0.5 for the counts in which \bar{r} exceeds 4.0. This discrepancy between methods (i) and (ii) may perhaps be due to 10 being too low a value of N for both methods to be accurate, or it may be due to a departure from the negative binomial form of distribution.

Thus, to sum up, there is clear evidence that k increases somewhat as m increases (an effect already noticed with *Myzus persicae* on potato plants [2]), and a suggestion that the form of distribution may perhaps depart from an exact negative binomial. In such an extensive series of counts, in which 940 experimental units were observed and almost 5,000 individuals (eggs) were counted, it is not surprising to find some contradiction of the simple hypothesis we started with. The same is to be expected with almost any kind of statistical material. Much attention has been given to investigating the validity of applying analysis of variance methods to yields in agricultural field experiments (without the question being entirely settled yet), and no such investigation of the validity in practice of the methods outlined in this paper has been undertaken. Our hypothesis, of negative binomial distributions with constant k , is the simplest we can make that is at all plausible; and the methods based on it are, if not elegant, at least not impossibly clumsy. It is suggested that no serious error will attend their use.

Accordingly, in further work on Price Jones's data, it would be reasonable to assume that k had a constant value of 0.5, if that facilitated the treatment. If we wished to correlate infestations at sites with other information about the sites, we could transform the total egg count per site, namely $10\bar{r}$, by the transformation

$$y = \text{Sinh}^{-1} \sqrt{\left(\frac{10\bar{r} + 0.375}{4.25} \right)},$$

and treat this as a normal variable with error variance $\frac{1}{4}\psi'(5) = 0.055$. In fact, no very interesting correlations were observed, as the information about the sites was rather imprecise; and whatever associations could be perceived, visually, from scatter diagrams, were equally clear when

untransformed counts were used. However, had such counts occurred in an experiment of the sort considered in §4, much clearer correlations would be expected; and the transformation would enable treatment effects to be investigated by analysis of variance.

REFERENCES

1. Anscombe, F. J. The transformation of Poisson, Binomial, and Negative Binomial Data. *Biometrika* 35, 246, 1948.
2. Anscombe, F. J. On Estimating the Population of Aphids in a Potato Field. *Annals of Applied Biology* 35, 567, 1948.
3. Beall, G. The Transformation of Data from Entomological Field Experiments so that the Analysis of Variance becomes Applicable. *Biometrika* 32, 243, 1942.
4. Davis, H. T. *Tables of the Higher Mathematical Functions* 2, Principia Press, Bloomington, 1935.
5. Fisher, R. A. The Negative Binomial Distribution. *Annals of Eugenics* 11, 182, 1941.
6. Fisher, R. A., Corbet, A. S., and Williams, C. B. The Relation between the Number of Individuals and the Number of Species in a Random Sample of an Animal Population. *Journal of Animal Ecology* 12, 42, 1943.
7. Kendall, D. G. On Some Modes of Population Growth Leading to R. A. Fisher's Logarithmic Series Distribution. *Biometrika* 35, 6, 1948.

QUERIES

66 QUERY: An experiment has been established to examine the influence of planted forest cover on a number of soil characteristics. Three kinds of trees have been grown on plots forming a randomized-block study with four replications. Because the individual plots in each block have to be rather large, the soil characteristics are to be obtained by stratified random sampling; and it is desired to obtain estimates of sampling errors. For these purposes each plot has been divided into 5 strata, within each of which 2 soil samples will be taken at random.

The resulting 120 samples are to be analyzed for a number of characteristics, such as organic matter and porosity. Because the laboratory work is costly, the workers would like to minimize it by pooling the field samples. In order to do this and still supply an estimate of sampling errors within strata, the following proposal has been made; can you tell us whether it is sound?

The 10 samples for each plot would be combined into two composite samples, each of which would contain one field sample drawn at random from the two in each of the five strata. Data obtained from the resulting 24 composite samples would be analyzed as follows:

<i>Source of Variation</i>	<i>D/F</i>
Treatments (<i>T</i>)	2
Blocks (<i>B</i>)	3
Experimental Error (<i>TB</i>)	6
Sampling Error	12

By this scheme no estimate of the variance among strata would be available; but this is of minor interest. The main reason for estimating the sampling error is to provide a basis for adjusting, if necessary, the number of future samples in the same strata. For this purpose the variance of a single soil sample within a stratum (s_x^2) would be estimated as $5s_r^2$.

ANSWER: The method you propose is good. You are really specifying two, randomly placed sampling units in each plot.

The five soil samples in each sampling unit are customarily taken in some systematic pattern (analogous to the knight's move, for example) with a randomly chosen starting point. The scheme you outline is one way of avoiding the possibility that the two units lie closely parallel.

QUERY: I have some data consisting of measurements of three
67 cord properties (X_1 = strength, X_2 = size, and X_3 = moisture regain) for each of five samples in each of three groups representing three somewhat different manufacturing processes and raw material but which are not known necessarily to affect the above cord properties.

The object was to determine the best estimate obtainable from these data of the coefficient of regression of X_1 on X_3 with statistical control of X_2 , rather than any significance of differences between groups for any of the properties. The coefficients determined from the pooled square and cross products of deviations from group means were outside the range of coefficients determined from individual groups as follows and the variance analysis did not indicate the regression to differ significantly between groups.

Regressions	Group I	Group II	Group III	Pooled
$b_{12 \cdot 3}$	3 932	4 206	6 460	507
$b_{13 \cdot 2}$.019	.023	.186	.012

At first, my interest was more the mathematical one of whether it could be proved that such values were or were not computationally possible. Since it has occurred with other data, and each time the computations were carefully checked, my question is now more the statistical one of whether the pooled results produce the best estimates of the true regression coefficient in such cases.

These data were not derived from a planned experiment but were an attempt to analyse available information as a guide to further study.

The situation presented is as unusual in my experience as
ANSWER: it evidently was in yours. As you have decided, the values you got are not precluded in the algebraic setup. I suspect such relationships would not often occur in samples of more usual sizes. Your degrees of freedom are so few that the tests of homogeneity cannot be expected to detect even large departures from hypothetical conditions.

Since your data are inadequate to furnish desired evidence, your question as to which regression produces the best estimate of the population coefficient must be answered (if there is any answer) by other knowledge you may have about the sampled population. If these were experimental data instead of fictitious, you might know, for example, that the manufacturing processes were properly controlled for the produc-

tion of uniform quality and that deviations from the regressions may be considered homogeneous. If homogeneity is reasonably assumed, you would then have to decide whether the three manufacturing processes may be considered to affect or not to affect the regressions—the statistical evidence is suspect. If you believe that the regression coefficients are really drawn from a common population, then the pooled coefficients are the ones to use. But if you have reason to believe that the three manufacturing processes may affect the parameters, then the individual regressions should be used for each process.

If you do not have the knowledge necessary to make these decisions, then the only way to proceed is to increase your sampling sufficiently to get the information from the experimental data themselves.

QUERY: A problem has arisen in the course of my research for
68 which I have been unable to find an adequate answer. The experiment involves the determination of the effect of androgen treatment on the activity of the thyroid gland of male white leghorn chicks. The activity is correlated with the size of the cells in the gland hence 50 cells in each of 5 glands from each series (experimental and control) were measured.

If I am not wrong in my understanding of the problem three analyses of variance are open for possible consideration. (1) Using the sum of squares of treatments as related to the sum of squares of the individual cell measurements in estimating F , a highly significant value (47.5) results. (2) If F is calculated from the sum of squares of treatments and the sum of squares between thyroids then the result (3.89) suggests no difference between the populations. (3) Analysis of Variance using the means of cell measurements yields an F value (6.08) which indicates significance at about the 4% level.

The analyses of variance are as follows: (see next page)

I had decided that the first method would give me the most sensitive test. Using either of the other methods one does not take into account the variability within the thyroids which I believe is important in the analysis. Or can one ignore this variability and assume the means are true means since such a large number of measurements is taken in each gland? Yet, this does not seem legitimate to me. I am not interested in the effect of the treatment upon the *cells* of the thyroid but rather upon the *thyroid* as shown by the change in the cells. Yet the variability of cell heights should play some part in the analysis, I believe.

	D.F.	Ssq	\bar{x} Ssq
Between thyroids	9	75.90	8.43
Between cells	490	338.24	0.690
Treatments	1	32.77	32.77

$$(1) F = 32.77/0.690 = 47.5. \quad (2) F = 32.77/8.43 = 3.89$$

	D.F.	Ssq	\bar{x} Ssq
Between thyroids	8	0.8626	0.1078
Treatments	1	0.6554	0.6554

$$F = 0.6554/0.1078 = 6.08$$

I have been unable to find a method of setting the fiducial limits of the difference between the mean of the means of the two series, 0.52.

ANSWER: It seems more convenient to analyze variance not in two or three tables but in a single one, as follows:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square
Treatments	1	32.77	32.77
Thyroids, same Treatment	8	43.13	5.39
Cells, same Thyroid	490	338.24	0.690
Total	499	414.14	

In one of your tables, you include the single degree of freedom for treatments among the 9 for all thyroids. I have separated these degrees of freedom in the combined analysis.

In your analysis of means, the mean squares for thyroids and treatments are each 1/50 of the corresponding mean squares in my table.

It is clear that there is a great deal of variation among thyroids over and above that which can be accounted for by the sampling variation of cells from the same gland. In testing the hypothesis that the treatment is without effect on the thyroid glands of male white leghorn chicks, the real experimental error must include this variation among chicks. The variation among cells within thyroids is, in fact, a part of

the mean square for thyroids—you are right in thinking that it cannot be ignored. The test is,

$$F = 32.77 \cdot 5.39 = 6.08, \quad df = 1 \text{ and } 8, \quad P = 0.04$$

It is assumed that the chicks were taken at random from the sampled population of white leghorn males and that the cells which were measured constitute random selections from the cells of the several thyroids.

Another assumption which has been made above is that the thyroid means are normally distributed. If σ is the same for the treated and untreated populations, confidence limits may be set on the mean difference by use of $s_d = \sqrt{2(5.39)/250} = 0.207$ with $df = 8$. For $P = 0.95$, $t = 2.306$, so that the half interval is 0.48.

QUERY: The setup of an experiment was as follows: Six rations were compared using a total of 186 turkey poults. The birds were weighed and listed in order of weight. The 6 heaviest birds were assigned to rations 1, 2, 3, 4, 5, 6 respectively, the next six birds to rations 6, 5, 4, 3, 2, 1 respectively. This method of selection was continued until all 186 birds had been assigned to the 6 rations. The object of this procedure was, of course, to have all 6 groups start out the experiment with approximately the same average weight. At 11 weeks of age all birds were weighed and the weights subjected to an analysis of co-variance. Analysis of co-variance was used because although the average starting weight was approximately the same for each of the 6 groups there were some differences.

My contention has been that, apart from any question of distribution of sexes or need for replication of pens (in this experiment there was a single pen for each of the 6 diets and the males and females were analyzed separately) that *no accurate* estimate of the error of the group means is possible because the birds have not been assigned to the groups at random.

I feel that a fundamental and important point may be involved here since one very frequently finds described in the literature the statistical analysis of data from experiments in which the experimental animals have been assigned to the groups in a systematic or non-random manner.

ANSWER: The most serious flaw in the experiment you describe is the absence of true replicates. With a single pen for each of the six diets, there is no reliable estimate of the experimental error. One could compute the difference between cages and

perhaps a standard error of the difference, but there is no information as to how much of the difference can be attributed to diet and how much to other environmental factors.

Only one estimate of error occurs to me in this case. If the six treatments represent six different levels in a single dietary constituent such that the results can be fitted with a straight line, three degrees of freedom representing the variation around the fitted line would be available for an error term. If this were of the same magnitude as the variation among birds within cages, the latter might be given some credence as an estimate of experimental error.

The systematic assignment of poults in order of weight might be a source of trouble. Intuitively it would not worry me too much for I suspect that other factors may be more important than the variation within weight groups of six poults. A preferable scheme, of course, would be to assign the six birds in each successive weight-class to the rations at random or, preferably, in accord with the rows in a series of randomized 6×6 Latin squares. It is not too important that the concomitant or initial characteristic in covariance be assigned at random, provided that this does not bias the distribution to different treatments. This follows from the fact that covariance is basically a regression technique and the independent variate in regression can be selected arbitrarily by the experimenter without biasing the results.

In young animals, the use of initial weight as a basis of assignment to blocks, or as a covariate, is not very effective as a means of controlling error. But a knowledge of food consumed by the experimental units (single birds or small groups of them) often results in worthwhile gains in efficiency. This is an added argument in favor of true replication.

C. I. BLISS

THE BIOMETRIC SOCIETY

One of the first decisions of the Biometric Society was to apply for affiliation with the International Statistical Institute as "an international organization concerned with a field of statistical specialization." We are happy to report that this affiliation has now been completed. There will be an exchange of representatives between the two organizations and all members of the Society attending the Conference in Geneva will receive an invitation to the ISI meetings in Berne on September 4-10. Members of the ISI in turn are invited to attend the Second Biometric Conference in Geneva on August 30 to September 2, announced in preceding issues of *Biometrics*.

In order to increase our effectiveness internationally, we have been in contact with UNESCO through Professor Pierre Auger, Director of the Department of Natural Sciences, Professor P. Vayssière, Secretary-General of the International Union of Biological Sciences, Professor Stuart Mudd, Secretary of the IUBS, and others. Biometry concerns so many different sciences that no simple solution was immediately available. One reason is that our Society is organized with the individual member as the unit, instead of with the nation as the unit as in the international unions. Recently, the International Union of Biological Sciences has invited the Biometric Society to serve as a specialized Section of the IUBS, and this invitation has been accepted by the Council of the Society. As a Section, \$200 has been allotted by the IUBS for expenses in 1949 and funds to aid in the publication of the proceedings of the Second International Biometric Conference at Geneva have been included in their budget for 1950. An International Union of Mathematical Sciences is projected and when formed, it is proposed that a mixed Commission on biometry under the International Council of Scientific Unions should take over the functions filled now by the new section of the IUBS.

As a result of the recent balloting, D. F. Votaw and E. K. Harris, tellers for Society, announce the adoption of the two amendments to the Constitution and the election of the following Council members for the term 1949-51 inclusive: J. Berkson, W. G. Cochran, D. Mainland, V. G. Panse, O. E. Sette and F. Yates.

As this issue of *Biometrics* went to press, the two new regions of the Society were proceeding with their organization. The Région Française,

comprising 45 members, held its first formal meeting on March 15 at the Laboratoire de Zoologie de la Sorbonne. By-laws for the Region were considered. It was agreed that the Région Française should form an official French society conforming to the French law of 1901 governing such associations. Draft by-laws will be submitted to all regional members and voted upon at the next meeting scheduled in May. The meeting named a regional committee of three, consisting of Mlle. Colette Rothschild and M. Lamotte (both of the Centre National de la Recherche Scientifique, Paris) and Dr. Marcel P. Schutzenberger, of the Hôpital St. Louis. Two papers were presented at this first meeting, one by Dr. Leon Vaugien entitled "Poids relatifs de la thyroïde, des surrénales, et de l'hypophyse antérieure chez les oiseaux" and another, "Analyse de la relation entre période d'incubation et nombre de particules virulentes injectées, dans le cas de la sensibilité héréditaire au gaz carbonique chez la *Drosophile*" by Professor Philippe L'Heritier and M. Kriatchko.

As reported in our last issue, the first meeting of the Indian Region was held in Allahabad on January 5, 1949. The following committee was elected to complete the regional organization: Professor P. C. Mahalanobis, Dr. U. S. Nair, Dr. P. V. Sukhatme, Dr. R. C. Bose, Dr. B. Ramamurthy, Dr. C. R. Rao and Mr. V. M. Dandekar. Draft by-laws have been prepared and sent to some 40 members of the Indian Region for approval. They provide for a regional vice-president, a secretary, a treasurer and a regional committee of nine members, all of whom will be voted upon by mail ballot together with the by-laws. Through Vice-President Mahalanobis, the Indian Statistical Institute has offered facilities for housing the regional office. This will aid in expanding the scope and activities of the Indian Region.

On April 20 the Eastern North American Region sponsored a joint session with the American Society for Pharmacology and Experimental Therapeutics at the Detroit meeting of the Federation of American Societies for Experimental Biology. The program consisted of a biometrical clinic on pharmacological problems. More than sixty questions had been submitted in advance by the pharmacologists, but only a small fraction of these could be considered by the panel consisting of C. I. Bliss, A. E. Brandt, K. A. Brownlee, S. Lee Crump, D. B. DeLury and Lloyd C. Miller (Chairman). About 150 attended the meeting.

The by-laws of the Western North American Region, adopted at Seattle, Washington, in November 1948, have been ratified by the Council of the Society and are reprinted below:

BY-LAWS—WESTERN NORTH AMERICAN REGION

The Western North American Region is governed by the constitution of the Biometric Society and the following regional By-laws.

1. The *aim* of the Region shall be to promote the understanding of quantitative biology and the application of statistical methods to biology.

2. *Membership.* By definition of the parent society the WNAR includes Mexico and those portions of United States and Canada lying west of approximately 104° West Longitude. All scientists residing in this region who have a substantial interest in quantitative biology, whether primarily biologists, statisticians or mathematicians, will be welcomed into the organization.

3. *Regional Committee.* The Regional Committee shall have authority to transact necessary business at all times when the annual meeting is not in session. It shall report to and be responsible to the membership as represented by the annual meeting and to the council of the parent society. It shall consist of the regional vice president, who shall be the presiding officer, the regional secretary treasurer and six ordinary members. The regional vice president and the regional secretary-treasurer shall be elected annually and may not serve for more than two consecutive terms, the ordinary members shall serve for three years, two to be elected each year. At the initial election six ordinary members shall be elected; these shall be divided by lot into three groups of two each, one to serve three years, another two years and a third one year.

Affiliations. The Regional Committee may affiliate with national societies for the purpose of joint meetings when common aims will be so served.

Annual meeting. There shall be an annual meeting of the region at a time and place to be determined by the Regional Committee on advice of members.

Amendment. These By-laws may be amended by a two-thirds vote of the members present at any annual meeting.

NEWS AND NOTES

ENGLAND—The University of Cambridge has set up a Statistical Laboratory under the auspices of the Faculty of Mathematics which was opened on March 2, 1949. The staff consists of **John Wishart**, Reader in Statistics since 1931 (who also has responsibility for the instruction of agricultural graduate students in statistics and field experimentation); **Henry E. Daniels** and **Frank J. Anscombe**, University Lecturers in Mathematics; and **Dennis V. Lindley**, University Demonstrator. The Laboratory accommodates the staff, graduate and visiting students, and computing assistants. A course is given yearly leading to a graduate Diploma in Mathematical Statistics, during which the candidates do work in one of a number of possible fields of application of statistical methods. The remainder of the graduates work for the PL.D. degree. The laboratory also offers a consultant service in statistics to University and other Departments, and is closely associated, in particular, with the University's Department of Applied Economics, directed by **Richard Stone**.

INDIA—**K. B. Madhava** returned from Government work in the Labour Bureau to The University Mysore from which he has retired. He writes, "I have not settled down to anything particularly yet, but I shall probably work on my own, combining my actuarial practice among my old insurance clients with some consulting statistical work." His new address is 70-A Stock Exchange Building, Apollo Street, Fort, Bombay. We would like to present a paragraph Mr. Madhava wrote on the scope of statistics from his article on "The rule of statistics in the formulation of a progressive labour policy." He says, "The value, rather the need, of statistics in practically every field of human endeavour in a present day administration of a world-knit progressive State is too well known to call for restatement at any length. Statistics has been described variously as the straw out of which bricks are made, as the brain and brawn of a Government, as the counterpart of operational research in relation to fighting services, etc. In essence, statistics may be likened to the all-pervasive science of meteorology, wide in coverage, dominating as a watch tower, and valuable in the service of man to navigate safely." We anticipate many active years of service are in store for Mr. Madhava. . . . The Indian Society of Agricultural Statistics

was founded in 1947 to promote the study of statistical theory and its application to agriculture, animal husbandry, agricultural economics and allied subjects. The first issue of the society's journal has appeared. Included are the Presidential Address by **The Honorable Rajendra Prasad**, Minister for Food and Agriculture, and three addresses given at a symposium on "Statistical Organization for India with special reference to Agriculture" by **The Honorable P. K. Shanmukham Chetty**, Minister for Finance, by **V. G. Panse**, Institute of Plant Industry, Indore and by **W. P. Natu**, Economic and Statistical Adviser, Government of India. The other articles were "Crop survey's in India" by **V. G. Panse** and **P. V. Sukhatme**, "A new approach to sampling distributions of the multivariate normal theory" and "On the distribution of estimated error components in analysis of variance and covariance" by **R. D. Narain**, "Estimation of genetic variability in plants" by **V. G. Panse** and **S. D. Bokil**, and "On fractional replication of the general symmetrical factorial design" by **K. Kishen**.

IRELAND—We are going to quote from a letter to the Secretary, The Biometric Society from **J. J. Brady**, Clontarf, Dublin. "You invite suggestions in regard to furthering the growth and development of The Biometric Society. My opinion is that the great majority of the articles published in *Biometrics* are too mathematical and theoretical to be of assistance to the vast numbers of research workers who require to use statistical methods in their experimental work but who have not the mathematical training essential to an understanding of such articles. . . . I think there is a great future for a statistical journal which would cater to the needs of the biological worker. Why not make *Biometrics* serve that purpose?" The Editorial Committee would like to have more articles which show the applications of statistical methods, or articles that combine applications with new point in theory. A large percentage of our subscribers are research workers in biological fields who want to learn more about these statistical tools.

JAPAN—A Japanese Conference on Experimental Statistics sponsored by the Ministry of Agriculture and Forestry of the Japanese government was held in Tokyo in March. About 75 experiment station representatives from throughout Japan attended this one-week conference. **Matayoshi Hatamura** of the Ministry's Agricultural Improvement Bureau served as chairman. **Warren H. Leonard**, chief of the Agriculture Division in the Occupation, discussed general field plot techniques.

Joseph C. Dodson, statistician in the Agriculture Division, discussed the design and statistical analysis of experiments. Japanese officials plan to hold a similar conference at a later date as a means of improving research methods used in their experiment stations. Mr. Dodson's assignment is in the production branch of the Agriculture Division. He is working on food collections. Mr. Leonard writes that he plans to return in August to the Colorado Agricultural College, Boulder.

NEW ZEALAND—J. T. Campbell, Senior Lecturer in Mathematics, Victoria University College, Wellington, wrote "When I returned to New Zealand in 1935, I found that there was little provision for instruction in statistical work. . . . The situation is improving." We would like to hear about the "use of twins in dairy cattle nutrition experiments and similar investigations". . . . At our request, **A. A. Rayner**, Biometrician, Extension Division, Department of Agriculture, Wellington, has written about his work. "The bulk of my work consists of the analysis of data from trials supervised by the Crop Experimentalist, **P. B. Lynch**, who is well known for his paper on the measurement of pastoral production. In 1947-48 there were 872 trials laid down mostly on the land of co-operating farmers. The trials embrace almost every type of crop and pasture grown by farmers in New Zealand, and many special types of investigation. I have little to do with routine analysis, but special problems are constantly being met, and there is always the designing of experiments. In the past few years the designs have been tending to increase in complexity. There is some work to be done for other Divisions of the Department, notably the Animal Research Division, whose milking-machine experiments are mainly the concern of **Jean Miller**. Horticulture has brought us experiments on storage of onions and packing of apples. For the Livestock Division, I have designed a national sampling survey for the assessment of such factors as pig losses and size of litters. I should say that the chief feature of our trials on farmer's land is the way in which they are laid down and harvested in accordance with normal farming practice, with modifications to suit the small experimental plots. For instance cereal trials are drilled with special 7-coulter drills and fertilizers are usually drilled with the seed, though this has its difficulties in factorial experiments. Sampling of plots at harvest has largely been abandoned in favour of header-harvesting. I think you call this sort of machine a 'combine'. A technique has been evolved in which the header is continuously in motion, and border rows are harvested at the same time, but discarded so far as weighing is concerned."

UNITED STATES—"At the national meeting of the American Home Economics Association in Minneapolis last June, the Research Departments Committee on Research Training, under the chairmanship of **Dr. Margaret A. Ohlson**, was commissioned to plan a workshop on design of experiments and surveys for persons directing and performing research in the field of Home Economics. Iowa State College was selected as the place to hold the workshop. The dates are June 13-25th, 1949. **Mr. Paul Homeyer**, Associate Professor in the Statistical Laboratory, has been designated as the Statistical Director of the workshop." . . . **T. W. Anderson**, Professor of Mathematical Statistics, Columbia University, New York City was a guest lecturer at Iowa State College, Ames during his Easter vacation. He gave two lectures on "Applications of multivariate analysis to problems in psychology and education" and a lecture on "Estimation of linear restrictions on regression coefficients and applications to econometric models". . . . We cannot keep up with **Geoffrey Beall** who is now with Research Laboratories, Swift and Company, Chicago. . . . When we reported last year about **K. A. Brownlee**, he was with Research and Development Department, The Distillers Company, Ltd., Surrey, England. Now, he is with E. R. Squibb and Sons, New Brunswick, New Jersey. . . . **S. Lee Crump**, has been at the University of Rochester, Rochester, New York since last summer. He went to New York from the Statistical Laboratory, Iowa State College, Ames. . . . Also to leave Ames last fall was **Walter T. Federer**, who is Professor of Biological Statistics, New York State College of Agriculture at Cornell University, Ithaca. . . . **David W. Fassett**, formerly at Cardiology Service, Jackson Memorial Hospital, Miami, Florida now gives this address: Laboratory of Industrial Medicine, Kodak Park Works, Eastman Kodak Company, Rochester. . . . **Marguerite F. Hall** is now with School of Medicine, Health Center, University of Washington, Seattle. She was an Associate Professor, School of Public Health, University of Michigan before going to the West coast. . . . **William P. Martin** is with the Agronomy Department, Ohio State University, Columbus. Previously, he was with the Southwest Forest and Range Experiment Station, Tucson, Arizona. . . . **Donald W. Maclaury**, Department of Poultry Husbandry, University of Kentucky, Lexington was formerly with the Department of Poultry Husbandry at Iowa State College, Ames. . . . **Maurice Whittinghall** of the Department of Zoology, University of North Carolina, Chapel Hill joined the Biology Division, Oak Ridge National Laboratories, for a six months period of research. He is doing research on a genetic problem using *Drosophila* as the experimental animal. . . . **J. A. Rafferty**, Captain M.C., will continue in his position as Chief of the Department of Biometrics, School of Aviation Medicine,

Randolph Field, Texas, when he becomes a civilian next month. Mr. Rafferty's concept of the functions of a department of biometrics is the full breadth of the subtitle on The Biometric Society's program announcements, "the mathematical and statistical aspects of biology." In line with his contention that "statisticians should get out of their rut of testing statistical hypotheses", his program is to include theoretical biology and medicine as well as traditional mass-data biometry and modern experimental statistics. At present the Department of Biometrics consists of forty personnel, mostly technical clerks, as computers and IBM operators. Mass-data projects on Air Force medical statistics are conducted as of special interest to the military establishment. About one-third of the workload is devoted to the design and analysis of laboratory and questionnaire investigations, done in consultation with research workers in all other departments of the School of Aviation Medicine. Another third of the workload is involved in mathematical and applied research in statistics, depending on the interests and capabilities of the professional statisticians and on the problems arising in applying statistics to aviation medicine. For instance, in the department, empirical sampling projects are under way concerning the relaxation of assumptions in the analysis of variance. Due to the importance of multivariate analysis in medical research, contracts for mathematical statistical research have been let to the University of California for work on discriminatory analysis, under the direction of **J. Neyman**; and to Yale University, for work in compound symmetry tests under the direction of **David Votaw**. Dr. Rafferty, as a medical research theorist, is interested in gathering into the Department of Biometrics as colleagues, mathematical statisticians and biomathematicians to work on mathematical models for biological and medical phenomena, to offer "cradle to grave" theory service to the experiments in the various basic medical research fields. . . . **Allyn W. Kimball, Jr.** has been with the Department of Biometrics in the capacity of experimental statistician since May, 1948. Mr. Kimball, a candidate for the Ph.D. degree in experimental statistics at the University of North Carolina, devotes much of his time to consultations with research workers in other departments on problems of design of experiments and interpretation of results. The progress made toward building up respect and confidence for statistics among the medical research personnel has been substantial and gratifying. In addition to these duties, Mr. Kimball conducts purely statistical research pertinent to aviation medicine and acts as project officer on contracts for statistical research awarded to civilian establishments.

B I O M E T R I C S

**The Biometrics Section of the
American Statistical Association**

TABLE OF CONTENTS

Optimum Allocation and Variance Components in Nested Sampling with an Application to Chemical Analysis	SOPHIE MARCUSE	189
Fitting a Straight Line when Both Variables are Subject to Error	M. S. BARTLETT	207
Relationships of Catch to Changes in Population Size of New England Haddock	HOWARD A. SCHUCK	213
One Degree of Freedom for Non-Additivity . . .	JOHN W. TUKEY	232
On a Statistical Approximation to the Infection Interval	J. B. CHASSAN	243
Queries		250
The Biometric Society		254
News and Notes		256

Material for *Biometrics* should be addressed to the Chairman of the Editorial Board, Institute of Statistics, North Carolina State College, Raleigh, N. C.; and material for Queries should go to "Queries", Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.

Officers

American Statistical Association

President: Simon Kuznets; President-Elect: S. S. Wilks; Vice-Presidents: L. S. Kellogg, H. A. Freeman, D. S. Brady; Secretary-Treasurer: M. M. Flood; Directors: C. H. Goulden, Isador Lubin, L. J. Reed, F. F. Stephan, W. L. Thorp, L. L. Thurstone.

Officers

Biometrics Section

Chairman: Margaret Merrell; Secretary: Max A. Woodbury; Section Committee: Joseph Berkson, Harold F. Dorn, Alton S. Householder, Jerzy Neyman and Joseph Zubin.

Editorial Board

Biometrics

Chairman: Gertrude M. Cox; Members: C. I. Bliss, D. J. Finney, H. C. Fryer, O. Kempthorne, A. M. Mood, Horace Norton, G. W. Snedecor and Jane Worcester.

Membership dues in the American Statistical Association are \$8.00 a year, of which \$4.00 is for a year's subscription to the quarterly *Journal*, \$.75 is for a year's subscription to *The American Statistician*. Members of the Association may subscribe to *Biometrics* at a special rate of \$2.00. Dues for associate members of the Biometrics Section are \$3.00 a year, of which \$2.00 is for a year's subscription to *Biometrics*. Single copies of *Biometrics* are \$1.00 each and annual subscriptions are \$3.00. Subscriptions and applications for membership should be sent to the American Statistical Association, 1603 K Street, N.W., Washington 6, D. C.

Entered as second-class matter, May 25, 1945, at the post office at Washington, D.C., under the Act of March 3, 1879. *Biometrics* is published four times a year—in March, June, September and December—by the American Statistical Association for its Biometrics Section. Editorial Office: 1603 K Street, N.W., Washington 6, D. C.

OPTIMUM ALLOCATION AND VARIANCE COMPONENTS IN NESTED SAMPLING WITH AN APPLICATION TO CHEMICAL ANALYSIS

SOPHIE MARCUSE

*U. S. Naval Research Laboratory,
Washington, D. C.*

INTRODUCTION

A SAMPLING TECHNIQUE frequently used in chemical and physical analyses for estimating the mean of a population is that of multiple random subsampling, called nested sampling by P. C. Mahalanobis.¹ For instance, when determining the moisture content of cheese, a food chemist might wish to select his samples randomly from different lots, and again from different cheeses of each lot, and finally make duplicate determinations on each cheese. A primary objective in the statistical design of such a sampling procedure is to minimize the cost of obtaining the sample estimate if the desired degree of precision is fixed, or conversely, to maximize the precision of the estimate obtained from a given amount of expenditure including personnel, time, and equipment. The question arises as to how the number of sampling units at each level should be determined to meet these optimum requirements assuming equal frequencies in the subclasses.

It is assumed in this paper that at each classification level, the cost is proportional to the number of units sampled at this level, and that the cost per sampling unit is known. Thus the total cost is a linear function of the numbers of sampling units at the various levels, with coefficients representing the (known) costs per sampling unit at these levels. On the other hand, the precision of the mean yielded by the experiment can be expressed in terms of the variance of this sample mean; it will then also be a linear function of the variances corresponding to each level, with coefficients involving the reciprocals of the number of units at the various levels. If the variances at the various levels are not known, they should be estimated from a preliminary experiment. The present paper discusses optimum allocation of the sampling units in nested sampling in terms of 3 levels. As an illustration of an experimental situation, a numerical example is given involving the estimation of variance components. In the appendix, the formulas for optimum allocation in nested sampling with k levels are derived.

¹For reference see M. Ganguli's paper on Nested Sampling [7].

For concreteness, we consider the above mentioned specific problem of planning in the most economical way an experiment in food chemistry designed to determine the moisture content of cheese, the subsampling levels involving lots, cheeses, and determinations. Clearly, the principles elucidated in terms of this particular problem for 3 levels are applicable to a wider class of problems involving more levels in subsampling, as, for instance, by expanding this simplified experiment to more than one factory. Also, they may be applied to other than chemical investigations involving nested sampling, for instance: in the determination of the breaking strength of a certain type of bronze, a metallurgist may wish to choose random samples from different ladles, then again from different molds of each ladle, and make duplicate determinations on the samples from each mold; in a manufacturing process, the subsampling categories may be lots, bags, and batches; in a gunnery experiment, test shooting may be done by different operators taking a number of observations on different runs; in agricultural investigations, the entire area under survey may be subdivided into a large number of zones, these in turn into a large number of smaller zones, and so on: in studies of spray deposit in insect work, plots, trees, and apple samples have been used as subsampling levels [2]. Examples of nested sampling in biological and industrial work together with analyses of variance components may be found in G. W. Snedecor's [10] and L. H. C. Tippett's [12] books. In designing a sample survey for estimating the jute crop in India, P. C. Mahalanobis [9] has used the cost function for considerations of optimum allocation and discussed their general application to large scale sample surveys; principles of optimum allocation in nested sampling have been used by M. H. Hansen et al. [8] in a sample survey of business involving 2-fold nested sampling from finite populations (countries, stores), and by L. H. C. Tippett [12] who describes an experiment where in obtaining soil samples from counts of cysts, a number of "borings" of soil were taken and then several counts made on each boring.

DEFINITION OF NESTED SAMPLING

The problem considered is one in which the total population is subdivided into primary sampling units (lots); these in turn are subdivided into secondary sampling units (cheeses) on which several measurements (determinations) are made representing the tertiary sampling units. The nested sample is obtained by selecting at random first n_1 primary (lots), then n_2 secondary (cheeses), and finally n_3 tertiary sampling units (determinations) from each of the preceding units, where n_1 , n_2 ,

n_3 represent the class frequencies. A measure of the variance of the sample mean in terms of the class frequencies is desired. Before deriving it, the structure of the mathematical model will be explained.

Let $x_{hi,}$ denote the j -th determination from the i -th cheese of the h -th lot. Assuming that the effects of the sampling units at the different levels are additive, we may describe an individual observation $x_{hi,}$ in nested sampling [7] as:

$$x_{hi,} = \mu + \xi_h + \eta_{hi} + \zeta_{hi,} \quad (1)$$

$h = 1, 2, \dots, n_1$ where h refers to the lot of cheese

$i = 1, 2, \dots, n_2$ where i refers to the cheese in each lot

$j = 1, 2, \dots, n_3$ where j refers to the determination on each cheese.

The value μ represents the general population mean and is thus a fixed constant. The components ξ_h , η_{hi} , $\zeta_{hi,}$ are random variables with means and covariances equal to zero and with variances equal to σ_1^2 , σ_2^2 , σ_3^2 , respectively, called variance components. Thus the components ξ_h , η_{hi} , $\zeta_{hi,}$ represent the effects peculiar to the lots, cheeses, and determinations, and the variance components the variabilities at the different levels.

VARIANCE OF SAMPLE MEAN AND ESTIMATION OF VARIANCE COMPONENTS IN NESTED SAMPLING

From the definition of an individual observation $x_{hi,}$ in nested sampling, given by equation (1), we have for the sample mean

$$\bar{x} = \mu + \frac{\sum_{h=1}^{n_1} \xi_h}{n_1} + \frac{\sum_{h=1}^{n_1} \sum_{i=1}^{n_2} \eta_{hi}}{n_1 n_2} + \frac{\sum_{h=1}^{n_1} \sum_{i=1}^{n_2} \sum_{j=1}^{n_3} \zeta_{hi,}}{n_1 n_2 n_3} \quad (2)$$

Then because of the assumptions made for the random variables ξ_h , η_{hi} , $\zeta_{hi,}$ we obtain for the variance of the sample mean

$$\sigma_{\bar{x}}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1 n_2} + \frac{\sigma_3^2}{n_1 n_2 n_3} \quad (3)$$

This expression gives the variance or precision of the sample mean as a linear function of the reciprocals of n_1 , $n_1 n_2$, and $n_1 n_2 n_3$ representing the total number of lots, cheeses, and determinations used. The coefficients are the variance components σ_1^2 , σ_2^2 , σ_3^2 , being the variances encountered at the 3 subsampling levels.

As long as the parameter values σ_1^2 , σ_2^2 , σ_3^2 are unknown, the variance function $\sigma_{\bar{x}}^2$ in (3) cannot be used for solving the problem to determine the optimum values of the class frequencies. On the other hand, if a set of class frequencies were given and used in performing an experiment in nested sampling, then the unknown parameters σ_1^2 , σ_2^2 , σ_3^2 could

be estimated from an analysis of variance of the experimental data. This dilemma² may be evaded by first carrying out a preliminary experiment in nested sampling³ using a set of arbitrarily chosen class

TABLE 1
ANALYSIS OF VARIANCE IN 3-FOLD NESTED SAMPLING

Source of Variation	Degrees of freedom	Mean Square	Expected Mean Square
Primary sampling units	$n_1^* - 1$	MS_1	$\sigma_3^2 + n_3^* \sigma_2^2 + n_3^* n_2^* \sigma_1^2$
Secondary sampling units within primary units	$n_1^* (n_2^* - 1)$	MS_2	$\sigma_3^2 + n_3^* \sigma_2^2$
Tertiary sampling units within secondary units	$n_1^* n_2^* (n_3^* - 1)$	MS_3	σ_3^2

frequencies. We will show how the data obtained from such a preliminary experiment give advance estimates of σ_1^2 , σ_2^2 , σ_3^2 , say s_1^2 , s_2^2 , s_3^2 , to be used for estimating the coefficients of the variance function.

Denote by n_1^* , n_2^* , n_3^* the given class frequencies of the preliminary experiment in nested sampling. Perform a customary analysis of variance on the observed data, as shown in the first 3 columns of table 1, where MS_1 , MS_2 , and MS_3 denote the mean squares corresponding to the primary, secondary, and tertiary sampling units. It can be shown that the expected values of the mean squares MS_1 , MS_2 , and MS_3 are the expressions shown in the last column of table 1⁴. Considering the estimates of these expressions by substituting the estimated variance components s_1^2 , s_2^2 , s_3^2 , we obtain the equations

$$\begin{aligned}
 MS_1 &= s_3^2 + n_3^* s_2^2 + n_3^* n_2^* s_1^2 \\
 MS_2 &= s_3^2 + n_3^* s_2^2 \\
 MS_3 &= s_3^2
 \end{aligned}
 \tag{4}$$

²See M. Friedman's discussion of a similar situation in planning an experiment ([11], p. 345).

³Or a mixed model design of experiment (e.g. randomized blocks or split plot) which includes the subsampling categories under consideration. Note that such a design might involve more degrees of freedom thus increasing the reliability of the estimated variance components ([3], [4]).

⁴Results for any number of sub-samplings and unequal frequencies are given by M. Ganguly [7].

Whence we have the solutions

$$\begin{aligned}s_3^2 &= MS_3 \\ s_2^2 &= \frac{MS_2 - MS_3}{n_3^*} \\ s_1^2 &= \frac{MS_1 - MS_2}{n_2^* n_3^*}\end{aligned}\tag{5}$$

in which the estimated variance components are expressed in terms of the mean squares calculated in the analysis of variance table of the experimental data from nested sampling.⁵ These equations can be extended from three to k subsamplings by the same reasoning.

OPTIMUM ALLOCATION IN 3-FOLD NESTED SAMPLING

The variance of the sample mean and the total cost expenditure for determining it, expressed in terms of the class frequencies, are the two functions needed for solving the optimum allocation problem under consideration. Considering the case of 3 levels, let $C(n_1, n_2, n_3)$ be the cost function and $V(n_1, n_2, n_3)$ the variance function, the variables n_1, n_2, n_3 representing the class frequencies. As given by equation (6), the cost function $C(n_1, n_2, n_3)$ is assumed to be an additive function of the costs at the three levels, that is the costs of n_1 primary, $n_1 n_2$ secondary, and $n_1 n_2 n_3$ tertiary sampling units altogether, the cost per primary, secondary, and tertiary sampling unit being c_1, c_2 , and c_3 respectively. The variance function $V(n_1, n_2, n_3)$ is given by equation (3) showing the variance of the sample mean, σ_s^2 , in 3-fold nested sampling; its parameters may be estimated from the data of a preliminary experiment by the analysis of variance procedure for estimating variance components as described above. Thus we have:

$$C(n_1, n_2, n_3) = c_1 n_1 + c_2 n_1 n_2 + c_3 n_1 n_2 n_3 \tag{6}$$

$$V(n_1, n_2, n_3) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_1 n_2} + \frac{\sigma_3^2}{n_1 n_2 n_3} \tag{3}$$

The problem of optimum allocation is to minimize $C(n_1, n_2, n_3)$ by proper choice of n_1, n_2, n_3 subject to the constraint that the allowable

⁵This analysis of the variance components was performed on data from nested sampling, which is a special case of Model II analysis of variance as shown below. If a similar analysis of variance components is routinely carried out on data belonging to Model I, the interpretation differs. In Model II, the computed variance components estimate the variances $\sigma_1^2, \sigma_2^2, \sigma_3^2$ associated with random factors, whereas in Model I, these are dummy symbols representing sums of squares of differences related to the variation of systematic (or fixed) factors ([1], [5]).

amount of variance is preassigned, say v , or to minimize $V(n_1, n_2, n_3)$ by proper choice of n_1, n_2, n_3 subject to the constraint that the total amount of cost is fixed, say c . Let n_{c1}, n_{c2}, n_{c3} and n_{v1}, n_{v2}, n_{v3} be the optimum solutions of the two problems respectively. By applying Lagrange multipliers it can be shown⁶ that these optimum values of n_1, n_2, n_3 are

$$\begin{aligned}
 n_{c1} &= \frac{\sigma_1}{v} \frac{\sum_{i=1}^3 (\sigma_i \sqrt{c_i})}{\sqrt{c_1}} \\
 n_{c2} &= \frac{\sigma_2}{\sigma_1} \sqrt{\frac{c_1}{c_2}} \\
 n_{c3} &= \frac{\sigma_3}{\sigma_2} \sqrt{\frac{c_2}{c_3}} \\
 n_{v1} &= \frac{\sigma_1}{\sum_{i=1}^3 (\sigma_i \sqrt{c_i})} \sqrt{c_1} \\
 n_{v2} &= \frac{\sigma_2}{\sigma_1} \sqrt{\frac{c_1}{c_2}} \\
 n_{v3} &= \frac{\sigma_3}{\sigma_2} \sqrt{\frac{c_2}{c_3}}
 \end{aligned} \tag{8}$$

The sets of equations (7) and (8) show similar features. Except for the first level, the optimum combination of the number of sampling units is independent of the given degree of precision or the fixed total cost, being the same whether the precision or the amount of cost is assigned beforehand. Therefore, when planning an experiment in nested sampling the analyst need be concerned with the given cost or precision only in selecting the number of primary sampling units. Clearly, an increase in funds would be utilized most efficiently, that is resulting in the highest possible precision, by a proportional increase in the number of primary sampling units, and similarly, the most economical way for attaining a higher degree of precision would consist in choosing a correspondingly greater number of primary sampling units.

In many instances, the research analyst might not wish to depend

⁶See appendix for development of these formulas.

on considerations of optimum allocation in the choice of the frequencies at all levels, but might prefer to take, for instance, duplicate or triplicate determinations from each cheese for check purposes, thus preassigning the class frequency associated to the tertiary sampling unit, n_3 . If n_3 is prefixed, the corresponding optimum allocation formulas⁷ are

$$n'_{c1} = \frac{\sigma_1}{\sqrt{c_1}} \frac{\sigma_1 \sqrt{c_1} + \sqrt{\left(\sigma_2^2 + \frac{\sigma_3^2}{n_3}\right)(c_2 + c_3 n_3)}}{\sqrt{c_1}} \quad (9)$$

$$n'_{c2} = \frac{c_1}{\sigma_1} \sqrt{\frac{c_1}{c_2 + c_3 n_3}}$$

in the case that the variance v is given; and

$$n'_{v1} = \frac{\sigma_1}{\sigma_1 \sqrt{c_1} + \sqrt{\left(\sigma_2^2 + \frac{\sigma_3^2}{n_3}\right)(c_2 + c_3 n_3)}} \frac{c}{\sqrt{c_1}} \quad (10)$$

$$n'_{v2} = \frac{\sqrt{\sigma_2^2 + \frac{\sigma_3^2}{n_3}}}{\sigma_1} \sqrt{c_2 + c_3 n_3}$$

in the case that the total cost c is given.

NUMERICAL EXAMPLE

The figures shown in table 2 are results from analyses of samples of cheese for the determination of moisture content.⁸ They will serve as the preliminary data for obtaining estimates of the variance components. The experimental set-up in nested sampling involves duplicate determinations made on 2 cheeses from each of 3 lots, the different cheeses and the different lots being randomly selected ($n_1^* = 3$, $n_2^* = 2$, $n_3^* = 2$).

The first 4 columns of table 3 show the results of an analysis of variance of these data. In nested sampling the sums of squares may be calculated as follows: Consider first table 2 (in which there are 3 factors: duplicates, cheeses, and lots) and refer to the figures, representing 1 determination, as "totals." Subsequently, obtain the totals

⁷See appendix for development of formulas in which all but the first K are fixed.

⁸The data are drawn from "Report on Sampling Fat and Moisture in Cheese" by William Horwitz and Lila F. Knudsen, J. Ass. Off. Agr. Chem., vol. 31 (1948), pp. 300-306; slight modifications have been made for illustrative purposes. The author acknowledges the suggestions of Lila F. Knudsen.

TABLE 2
MOISTURE CONTENT OF 2 CHEESES FROM EACH OF 3 DIFFERENT LOTS,
DETERMINED 2 TIMES

Cheese	Lot		
	I	II	III
1	39.02	35.74	37.02
	38.79	35.41	36.00
2	38.96	35.58	35.70
	39.01	35.52	36.04

for the duplicates on each cheese (there remain 2 factors: cheeses and lots), and also the totals of the 4 determinations on each lot (there remains 1 factor: lots), in addition to the total for the entire table (no

TABLE 3
ANALYSIS OF VARIANCE OF DATA ON MOISTURE CONTENT OF CHEESE
GIVEN IN TABLE 2

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Expected Mean Square	Estimated Variance Components
Lots	2	$SS_1 = 25.9001$	$MS_1 = 12.9501$	$\sigma_1^2 + 2\sigma_2^2 + 4\sigma_3^2$	$s_1^2 = 3.2028$
Cheeses	3	$SS_2 = 41.60$	$MS_2 = 13.87$	$\sigma_1^2 + 2\sigma_2^2$	$s_2^2 = 0.143$
Replications	6	$SS_3 = 0.620$	$MS_3 = 0.103$	σ_3^2	$s_3^2 = 0.103$

factor remains). Denote by Q_3 , Q_2 , Q_1 , and Q_0 the sum of squares of these corresponding totals divided by the number of determinations making up each total:

$$Q_3 = 39.02^2 + 38.79^2 + \dots + 35.70^2 + 36.04^2 = 16,365.5607$$

$$Q_2 = \frac{77.81^2 + 77.97^2 + 71.15^2 + 71.10^2 + 73.02^2 + 71.74^2}{2}$$

$$= 16,364.8988$$

$$Q_1 = \frac{155.78^2 + 142.25^2 + 144.76^2}{4} = 16,364.4821$$

$$Q_0 = \frac{442.79^2}{12} = 16,338.5820$$

Then the sums of squares in analysis of variance, SS_1 , SS_2 , SS_3 , are the successive differences of these expressions:

$$SS_1 = Q_1 - Q_0 = 25.9001$$

$$SS_2 = Q_2 - Q_1 = 0.4166$$

$$SS_3 = Q_3 - Q_2 = 0.6620^a$$

The sums of squares and the corresponding mean squares are shown in columns 3 and 4 of table 3. The estimated variance components s_1^2 , s_2^2 , s_3^2 , shown in the last column of table 3, follow from equations (5). These values represent the advance estimates from the preliminary data to be used in the planning of the experiment.

The problem of designing an experiment with optimum allocation may arise in chemical laboratory work, e.g., when it is desired to set up in the most economical way routine analyses of samples of cheese for the determination of moisture content. In the example under consideration we assume that the chemist wants to spend not more than 60 dollars altogether to be allocated in such a way that the highest precision results; that he requires duplicate determinations for check purposes; and that the cost factors per lot, cheese, and determination are 10, 3, and 1 dollar respectively. Since these requirements prefix the class frequency n_3 and the total cost C , formulas (10) are appropriate. Substituting $n_3 = 2$, $c = 60$, $c_1 = 10$, $c_2 = 3$, and $c_3 = 1$, and for the variances σ_1^2 , σ_2^2 , σ_3^2 their estimates $s_1^2 = 3.2028$, $s_2^2 = 0.0143$, $s_3^2 = 0.1103$, we obtain:

$$n'_{r1} = 5.43 \quad n'_{r2} = 0.21$$

The corresponding integer values have to be chosen in accordance with the conditions of the experiment. Since n_2 , the number of cheeses selected from each lot, must be at least one, the number of lots, n_1 , may be reduced. An examination of the integers smaller than n'_{r1} shows that $n_1 = 4$ together with $n_2 = 1$ fulfill the required conditions. Thus 4 lots and 1 cheese give the optimum solution for the problem under consideration.

The merit of this optimum combination may be judged by comparing it to other combinations of class frequencies. In table 4 a number of various combinations (columns 1 and 2) are presented together with the precision of the sample mean (columns 5 and 6) and

^aUsing the figures given for Q_1 , Q_2 above, we have $Q_3 - Q_2 = .6619$ instead of .6620. Such a difference in the last decimal place is due to rounding off results, intermediate computations being carried out to more decimal places.

TABLE 4

ESTIMATED PRECISION AND COST OF DETERMINING MOISTURE CONTENT OF CHEESE WHEN A SPECIFIED NUMBER OF LOTS (n_1) AND A SPECIFIED NUMBER OF CHEESES FROM EACH LOT (n_2) ARE USED AND TWO DETERMINATIONS ($r_1 = 2$) ARE MADE ON EACH CHEESE. CONSTANTS USED ARE ADVANCE ESTIMATES CALCULATED FROM PRELIMINARY DATA (TABLES 2 AND 3)

Formulas used:	Constants used:
$N = n_1 n_2 n_3$	$n_3 = 2$
$C = c_1 n_1 + c_2 n_1 n_2 + c_3 n_1 n_2 n_3$	$c_1 = 10, c_2 = 3, c_3 = 1$
$V = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_1 n_2} + \frac{s_3^2}{n_1 n_2 n_3}$	$s_1^2 = 3.2028, s_2^2 = .0143, s_3^2 = .1103$
$CV = \frac{\sqrt{v}}{\bar{x}} \times 100$	$\bar{x} = 36.90$

Number of—		Expenditure		Estimated Precision	
Lots	Cheeses	Number of Determinations	Total Cost in dollars	Variance of mean	Coefficient of Variation
n_1	n_2	N	C	V	CV
(1)	(2)	(3)	(4)	(5)	(6)
5	3	30	125	0.6452	2.18
5	2	20	100	0.6475	2.18
5	1	10	75	0.6544	2.19
4	3	24	100	0.8065	2.43
4	2	16	80	0.8094	2.44
4	1	8	60	0.8181	2.45
3	3	18	75	1.0753	2.81
3	2	12	60	1.0792	2.82
3	1	6	45	1.0907	2.83
2	3	12	50	1.6130	3.44
2	2	8	40	1.6188	3.45
2	1	4	30	1.6361	3.47
1	3	6	25	3.2259	4.87
1	2	4	20	3.2375	4.88
1	1	2	15	3.2722	4.90

the expenditure involved in determining it (columns 3 and 4). Column 3 shows the total number of determinations made, the total cost is given in column 4, and column 6 compares the relative precision of

the sample mean, indicated by its coefficient of variation, to the absolute precision in terms of the variance (column 5). Duplicate determinations are used throughout. It can be seen that the 4-1-2 combination is more economical than the 3-2-2 combination—the one used in the preliminary experiment—since it obtains a higher precision but requires the same cost (60 dollars). Also, the combination 3-2-2 is less efficient than the combination 3-1-2 since, for the same precision, the latter combination needs half the number of determinations and requires only 45 dollars instead of 60 dollars. In general, it pays to increase the number of lots instead of the number of cheeses since the former are more variable.

REMARKS ON NESTED SAMPLING AS A SPECIAL CASE OF
MODEL II ANALYSIS OF VARIANCE

The mathematical model of nested sampling as given by the fundamental equation (1) and its assumptions, is closely related to one specific mathematical model used in analysis of variance. Two models of analysis of variance, usually referred to as Model I and Model II, have been discussed recently by S. L. Crump [3] and C. Eisenhart [5]. It seems worthwhile to show that, in virtue of the underlying assumptions, nested sampling represents a special case of Model II of analysis of variance.

The two different models of analysis of variance involve the analysis of two different types of factors: systematic factors in Model I and random factors in Model II. A factor such as "treatment" or "lot" is a random or a systematic factor depending on the way its variants are chosen. Here the term "variant" of a factor is used based on Fisher's terminology [6], for instance, the variants of the factor "treatment" may be e.g. "nitrogen" and "phosphate" and different lots the variants of the factor "lot." When an experimenter selects the two treatments "nitrogen" and "phosphate," he selects them systematically from a population of possible treatments on the basis of subject matter judgment; on the other hand, when selecting different lots of material for studying the effects of the treatments, he generally bases his choice on random selection ([5], [10] Chapter 8). Since systematically chosen variants produce systematic variation and randomly chosen variants random variation, the type of factor may be determined according to the issue: systematic or random variation. Usually, "methods" and "treatments" represent systematic factors, "blocks" and "lots" random factors, whereas factors such as "days" or "animals" or "locations" may represent either systematic or random factors; both types of factor will often occur in the same experiment; then the model is a mixed one.

Now the factors encountered in nested sampling are the primary, secondary, tertiary sampling units (lots, cheeses, determinations). Under the assumptions made, the variants of these factors, i.e. the units selected at each level, were chosen randomly. These factors, therefore, are random factors and thus nested sampling belongs to Model II.

In order to describe more accurately the relationship of nested sampling to Model II of analysis of variance, we subdivide the random factors of Model II into two categories: cross classified¹⁰ with respect to another factor or not. For instance, in the 2 factor "day-animal" experiment discussed by C. Eisenhart [5] as an example of Model II, the random factor "animal" is cross classified with respect to the factor "days," each of the randomly chosen animals being tested on all days (the analysis of variance table contains: "Between days," "Between animals," and "Residual" with $d - 1$, and $a - 1$, and $(a - 1)(d - 1)$ degrees of freedom respectively). On the other hand, there would be no cross classification, if on each day a number of animals were randomly chosen for testing, as for instance in an inoculation experiment affecting the sensitivity of the animal (the analysis of variance contains: "Between days," and "Between animals within days" with $d - 1$, and $d(a - 1)$ degrees of freedom respectively). Likewise, no cross classification would be involved for the random factor "animal" if each animal would be tested on a couple of days which were randomly selected, as e.g. if only one animal could be tested per day (the analysis of variance contains: "Between animals," and "Between days within animals" with $(a - 1)$, and $a(d - 1)$ degrees of freedom respectively). Nested sampling represents the second category of Model II in which the random factors involved are not cross classified since for each primary sampling unit a number of secondary sampling units is selected randomly, and so on. The question as to which order of sub-sampling should be adopted in the nested sampling procedure, as, for instance, whether to use "animals" as primary sampling units and "days" as secondary sampling units, or conversely, is a decision to be made on the basis of subject matter judgment.

APPENDIX

We shall now derive the optimum values of the class frequencies, given for the three-fold level by formulas (7), (8), (9), and (10), for the general case of k -fold nested sampling. Instead of solving the prob-

¹⁰This term is not synonymous with "ordered". Note that items in table 2 below are ordered for purely designative reasons there being neither a cross classification nor an element of "sequence" involved.

lem directly by introducing the Lagrange multiplier, we will apply this procedure to a pair of generalized functions. We then obtain as special cases the solution formulas for optimum allocation in

- i. k -fold nested sampling
- ii. k -fold nested sampling in which some class frequencies are fixed beforehand
- iii. stratified sampling from finite population (k strata, 2 levels).

a. *Minimum Problem for 2 Generalized Functions*

Let the two generalized functions be

$$F_1(N_1, \dots, N_k) = \sum_{i=1}^k a_{1i} N_i + a_1 \quad (11)$$

$$F_2(N_1, \dots, N_k) = \sum_{i=1}^k \frac{a_{2i}}{N_i} + a_2 \quad (12)$$

where N_1, \dots, N_k denote variables and a_1, a_2 , and a_{1i}, a_{2i} ($i = 1, \dots, k$) are constants.

Consider first the problem to minimize $F_1(N_1, \dots, N_k)$ subject to the side condition

$$F_2(N_1, \dots, N_k) = b_2 \quad (13)$$

where b_2 is a constant. Using the Lagrange multiplier λ in the usual way, we let the derivatives of $F_1 + \lambda F_2$ with respect to N_i ($i = 1, \dots, k$) be zero, and obtain

$$a_{1i} - (\lambda a_{2i} / N_i^2) = 0$$

or

$$N_i = \sqrt{\lambda} \sqrt{a_{2i} / a_{1i}}$$

Substituting these values of N_i in (13), where F_2 is given by (12), we have

$$F_2(N_1, \dots, N_k) = (1/\sqrt{\lambda}) \sum_{i=1}^k \sqrt{a_{1i} a_{2i}} + a_2 = b_2$$

Therefore

$$\sqrt{\lambda} = \frac{\sum_{i=1}^k \sqrt{a_{1i} a_{2i}}}{b_2 - a_2}$$

Hence we obtain the optimum values

$$N_i = \frac{\sum_{i=1}^k \sqrt{a_{1i} a_{2i}}}{b_2 - a_2} \sqrt{\frac{a_{2i}}{a_{1i}}} \quad (14)$$

Similarly, we obtain the solution of the problem to minimize $F_2(N_1, \dots, N_k)$ subject to the side condition

$$F_1(N_1, \dots, N_k) = b_1 \quad (15)$$

where b_1 is a constant:

$$N_{21} = \frac{b_1 - a_1}{\sum_{i=1}^k \sqrt{a_{1i} a_{2i}}} \sqrt{\frac{a_{21}}{a_{11}}} \quad (16)$$

Now introduce the variables

$$n_1 = N_1, \quad n_i = N_i / N_{i-1} \quad (i = 2, \dots, k) \quad (17)$$

then $N = n_1 \dots n_i (i = 1, \dots, k)$. Substituting the new variables in (11) and (12), we obtain the functions

$$f_1(n_1, \dots, n_k) = \sum_{i=1}^k a_{1i} n_1 \dots n_i + a_1 \quad (18)$$

$$f_2(n_1, \dots, n_k) = \sum_{i=1}^k \frac{a_{2i}}{n_1 \dots n_i} + a_2 \quad (19)$$

Substituting (14) in (17), we find that the minimum solutions of $f_1(n_1, \dots, n_k)$ under the side condition $f_2(n_1, \dots, n_k) = b_2$ are:

$$n_{11} = \frac{\sum_{i=1}^k \sqrt{a_{1i} a_{2i}}}{b_2 - a_2} \sqrt{\frac{a_{21}}{a_{11}}} \quad (20)$$

and

$$n_{1i} = \sqrt{\frac{a_{2i} a_{1i-1}}{a_{1i} a_{2i-1}}} \quad (i = 2, \dots, k)$$

Similarly, substituting (16) in (17), we find the minimum solutions of $f_2(n_1, \dots, n_k)$ under the side condition $f_1(n_1, \dots, n_k) = b_1$:

$$n_{21} = \frac{b_1 - a_1}{\sum_{i=1}^k \sqrt{a_{1i} a_{2i}}} \sqrt{\frac{a_{21}}{a_{11}}} \quad (21)$$

and

$$n_{2i} = \sqrt{\frac{a_{2i} a_{1i-1}}{a_{1i} a_{2i-1}}} \quad (i = 2, \dots, k)$$

Note that $n_{1i} = n_{2i} (i = 2, \dots, k)$.

b. *Application to Optimum Allocation Problems in Sampling*i. *Nested Sampling*

Substituting $a_{1i} = c_i$, $a_{2i} = \sigma_i^2$ and $a_1 = a_2 = 0$ in (18) and (19), we obtain the 2 functions

$$g_1(n_1, \dots, n_k) = \sum_{i=1}^k c_i n_1 \cdots n_i \quad (22)$$

$$g_2(n_1, \dots, n_k) = \sum_{i=1}^k \frac{\sigma_i^2}{n_1 \cdots n_i} \quad (23)$$

These functions represent the general case of the cost function $C(n_1, n_2, n_3)$ and the variance function $V(n_1, n_2, n_3)$ used above in section 4. Setting $b_1 = c$ and $b_2 = v$ yields the corresponding side conditions. Therefore applying formulas (20) and (21), we have as the minimum solutions of $g_1(n_1, \dots, n_k)$ under the side condition $g_2(n_1, \dots, n_k) = v$

$$n_{11} = \frac{\sigma_1}{v} \frac{\sum_{i=1}^k (\sigma_i \sqrt{c_i})}{\sqrt{c_1}} \quad (24)$$

and

$$n_{1i} = \frac{\sigma_i}{\sigma_{i-1}} \sqrt{\frac{c_{i-1}}{c_i}} \quad (i = 2, \dots, k)$$

and as the minimum solutions of $g_2(n_1, \dots, n_k)$ under the side condition $g_1(n_1, \dots, n_k) = c$

$$n_{21} = \frac{\sigma_1}{\sum_{i=1}^k (\sigma_i \sqrt{c_i})} \frac{c}{\sqrt{c_1}} \quad (25)$$

and

$$n_{2i} = \frac{\sigma_i}{\sigma_{i-1}} \sqrt{\frac{c_{i-1}}{c_i}} \quad (i = 2, \dots, k)$$

Specializing equations (24) and (25) to the case $k = 3$ yields equations (7) and (8). Specializing equation (25) to the case $k = 2$ and letting cost be expressed in terms of time, $c_1 = kt$, $c_2 = t$, gives equation 10.32 in L. H. C. Tippett's book [12].

ii. *Nested Sampling with Some Prefixed Class Frequencies*

Let n'_1, \dots, n'_i be the unknown frequencies and $n_{k'+1}, \dots, n_k$ be

fixed beforehand. The equations (22) and (23) may then be rewritten in terms of $n'_1, \dots, n'_{k'}$ as follows:

$$\begin{aligned} h_1(n'_1, \dots, n'_{k'}) &= \sum_{i=1}^{k'} c_i n'_1 \cdots n'_i + n'_1 \cdots n'_{k'} \sum_{i=1}^{k'-1} c_{i+1} n'_{i+1} \cdots n'_{i+1} \\ &= \sum_{i=1}^{k'} c'_i n'_1 \cdots n'_i \end{aligned} \quad (26)$$

where $c'_i = c_i$ ($j = 1, \dots, k' - 1$)

and

$$c'_{k'} = c_{k'} + \sum_{i=1}^{k'-1} c_{i+1} n'_{i+1} \cdots n'_{i+1} \quad (27)$$

$$\begin{aligned} h_2(n'_1, \dots, n'_{k'}) &= \sum_{i=1}^{k'} \frac{\sigma_i^2}{n'_1 \cdots n'_i} + \frac{1}{n'_1 \cdots n'_{k'}} \sum_{i=1}^{k'-1} \frac{\sigma_{i+1}^2}{n'_{i+1} \cdots n'_{i+1}} \\ &= \sum_{i=1}^{k'} \frac{\sigma'^2_i}{n'_1 \cdots n'_i} \end{aligned} \quad (28)$$

where $\sigma'_i = \sigma_i$ ($j = 1, \dots, k' - 1$)

and

$$\sigma'^2_{k'} = \sigma_{k'}^2 + \sum_{i=1}^{k'-1} \frac{\sigma_{i+1}^2}{n'_{i+1} \cdots n'_{i+1}} \quad (29)$$

Thus the functions h_1 and h_2 of the variables $n'_1, \dots, n'_{k'}$, given by (26) and (28), represent the same types of function as the functions g_1 and g_2 of the variables n_1, \dots, n_k given by (22) and (23). Therefore the minimum solutions of $h_1(n'_1, \dots, n'_{k'})$ and $h_2(n'_1, \dots, n'_{k'})$ under the side conditions $h_2(n'_1, \dots, n'_{k'}) = v$ and $h_1(n'_1, \dots, n'_{k'}) = c$ respectively, may be obtained from equations (24) and (25) by replacing k by k' , σ by σ' , and c by c' , and then substituting back σ'_i and c'_i ($j = 1, \dots, k'$) from equations (27) and (29).

For $k = 3$, $k' = 2$ we obtain from (27) and (29)

$$c'_1 = c_1 \quad c'_2 = c_2 + c_3 n_3$$

$$\sigma'_1 = \sigma_1 \quad \sigma'^2_2 = \sigma_2^2 + \frac{\sigma_3^2}{n_3}$$

The substitution of these values into (24) and (25) after replacement of k, c, σ by k', c', σ' gives the formulas (9) and (10) used above.

Note that the results of b. ii. may also be obtained from a. and then b. i. be considered as the special case $k' = k$.

iii. *Stratified Sampling from Finite Populations*

We will indicate briefly the applicability of the above used generalized functions to stratified sampling involving two levels.

Let there be k strata in the population with M_i elements x_i in the i -th stratum ($i = 1, \dots, k; j = 1, \dots, M_i$). Assume that the N_i sample elements x_{ij} ($i = 1, \dots, k; j = 1, \dots, N_i$) are independently drawn at random from the k finite strata. Then the sample mean

$$\bar{x} = \frac{1}{M} \sum_{i=1}^k M_i \frac{\sum_{j=1}^{N_i} x_{ij}}{N_i}$$

has the variance

$$\sigma_x^2 = \frac{1}{M^2} \sum_{i=1}^k M_i^2 \frac{\sigma_i^2}{N_i} \frac{M_i - 1}{M_i - 1}$$

where $M = \sum_{i=1}^k M_i$ and σ_i^2 denotes the variance between elements in the i -th stratum. Thus we have

$$\sigma_x^2 = \sum_{i=1}^k \frac{a_{2i}}{N_i} + a_2$$

where $a_{2i} = \frac{M_i^3 \sigma_i^2}{M^2(M_i - 1)}$ and $a_2 = -\frac{1}{M^2} \sum_{i=1}^k \frac{M_i^2 \sigma_i^2}{M_i - 1}$

Let c_i be the cost per element in the i -th stratum and $c = \sum_{i=1}^k c_i N_i$ the total cost, then c may be written $c = \sum_{i=1}^k a_{1i} N_i + a_1$ where $a_{1i} = c_i$ and $a_1 = 0$. Thus c and σ_x^2 correspond to the functions $F_1(N_1, \dots, N_k)$ and $F_2(N_1, \dots, N_k)$ respectively in (11) and (12). Therefore equations (14) and (16) give the desired minimum solutions where b_1 and b_2 determine the side conditions corresponding to (13) and (15). In case the populations in the strata are large ($M_i \sim M_i - 1$), we obtain the well known optimum allocation formulas:

$$N_{1i} = \frac{\sum_{i=1}^k (M_i \sigma_i \sqrt{c_i})}{M^2 b_2 + \sum_{i=1}^k (M_i \sigma_i^2)} \frac{M_i \sigma_i}{\sqrt{c_i}}$$

$$N_{2i} = \frac{b_1}{\sum_{i=1}^k M_i \sigma_i \sqrt{c_i}} \frac{M_i \sigma_i}{\sqrt{c_i}}$$

LITERATURE CITED

- [1] Anderson, R. L. Use of Variance Components in the Analysis of Hog Prices in Two Markets, *J. Am. Stat. Ass.*, 42: 612-634, 1947.
- [2] Cassil, C. C., Wadley, F. M., and Dean, F. P. Sampling Studies on Orchard Spray Residues in the Pacific Northwest, *J. of Econ. Entom.*, 36: 227-231, 1943.
- [3] Crump, S. Lee. The Estimation of Variance Components in Analysis of Variance, *Biometrics Bulletin*, 2: 7-11, 1946.
- [4] Daniels, H. E. The Estimation of Components of Variance, *Supplement to the Journal of the Royal Statistical Society*, 6: 186-197, 1939.
- [5] Eisenhart, Churchill. The Assumptions Underlying the Analysis of Variance, *Biometrics Bulletin*, 3: 1-21, 1947.
- [6] Fisher, R. A. *The Design of Experiments*, 3rd Edition. Oliver and Boyd, Ltd., Edinburgh and London, 1942.
- [7] Ganguli, M. A Note on Nested Sampling, *Sankhya*, 5: 449-452, 1941.
- [8] Hansen, M. H., Hurwitz, W. N., and Gurney, M. Problems and Methods in a Sample Survey of Business, *J. Am. Stat. Ass.*, 41: 173-189, 1946.
- [9] Mahalanobis, P. C. On Large Scale Sample Surveys, *Philos. Transactions of the Royal Society, Series B, Biolog. Sciences*, 23: 329-451, 1944.
- [10] Snedecor, G. W. *Statistical Methods Applied to Experiments in Agriculture and Biology*, 4th Edition. The Collegiate Press, Inc., Ames, Iowa, 1946.
- [11] Statistical Research Group, Columbia University. *Selected Techniques of Statistical Analysis: For Scientific and Industrial Research and Production and Management Engineering*. McGraw-Hill Book Company, Inc., New York, 1947.
- [12] Tippett, L. H. C. *The Methods in Statistics*, 3rd Edition. Williams and Norgate, Ltd., London, 1940.

FITTING A STRAIGHT LINE WHEN BOTH VARIABLES ARE SUBJECT TO ERROR

M. S. BARTLETT

University of Manchester, England

INTRODUCTION

A SIMPLE METHOD of fitting a straight line when both variables are subject to error was examined by Wald (1) in 1940. The purpose of the present note is to present and illustrate a modification of Wald's method having the advantage in general of greater accuracy. Before any detailed exposition it will be as well to recall two important points:

- (i) a distinction must be made between the linear regression equation of a variable y on a second variable x , and a linear functional relation between two variables Y and X masked by errors. The former equation is still available for prediction even if the variable x is subject to error, but is not necessarily appropriate for a functional relation when one exists.
- (ii) it is possible to set up maximum likelihood equations for the second problem, but they do not lead to a unique solution without further assumptions, such as an assumption about the relative magnitude of the errors in x and y .

These points have been emphasized by many previous writers, for example, by Wald (1) or more recently by Lindley (2). In view of (ii) it is useful to consider, in the common case when the observations have equal weight, the following elementary method:

- (a) For the location of the fitted straight line use as one point the mean coordinates \bar{x} , \bar{y} , just as in the least-squares method.
- (b) For the slope, first divide the n plotted points into three groups, the equal numbers k in the two extreme groups being chosen to be as near $\frac{1}{3}n$ as possible (the three groups are non-overlapping when considered, say, in the x direction). The join of the mean coordinates \bar{x}_1 , \bar{y}_1 and \bar{x}_3 , \bar{y}_3 for the two extreme groups is used to determine the slope.

The only difference from Wald's original method is the use of three groups instead of two, for reasons which will be apparent from the results

of the next section.¹ It will also be shown that Wald's confidence interval method of assessing the accuracy (under suitable conditions) may be adapted to the present method.

EFFICIENCY IN A SPECIAL CASE

To get some idea of the efficiency of the method its accuracy is determined in a special case where least-squares is appropriate. It is assumed that observations y are available for $n = 2l + 1$ values $x \equiv X$ not subject to error and spaced at equidistant unit intervals. The least-squares estimate is known to provide the linear combination of the y 's providing an unbiased estimate of the true slope β in the functional relation

$$(1) \quad Y = \alpha + \beta X$$

with minimum variance when the differences $y - Y$ are uncorrelated and of constant variance σ^2 . The least-squares estimate

$$b = \sum y(x - \bar{x}) / \sum (x - \bar{x})^2$$

has error variance $\sigma^2 \sum (x - \bar{x})^2$, where $\sum (x - \bar{x})^2 = (\frac{1}{3})l(l + 1)(2l + 1)$ in the situation assumed in this section.

For comparison the error variance of the estimate

$$(2) \quad b' = \frac{\bar{y}_3 - \bar{y}_1}{\bar{x}_3 - \bar{x}_1}$$

of the last section is easily evaluated for any value of k . It is given by

$$\frac{2\sigma^2}{k(\bar{x}_3 - \bar{x}_1)^2} = \frac{2\sigma^2}{k(2l - k + 1)^2}.$$

The relative efficiency of b' is thus

$$(3) \quad E = \frac{3k(2l - k + 1)^2}{2l(l + 1)(2l + 1)}.$$

This is a maximum when

$$(2l - k + 1)(2l + 1 - 3k) = 0$$

with relevant root $k = (\frac{1}{3})(2l + 1) = \frac{1}{3}n$.

I am indebted to Professor Gerhard Fintner for drawing my attention to a previous discussion of this problem, with a similar conclusion, by Nair and Shrivastava (4) (see also Nair and Banerjee (5)). It might be noted that these authors propose using the two extreme groups out of three for location as well as slope, but recommendation (a) above is theoretically preferable. In the first of these two papers the extension of the method to fitting higher-order curves is also considered, though the optimum efficiency is not so high in such cases.

We then have

$$(4) \quad E = \frac{8(l + \frac{1}{2})^2}{9l(l + 1)} \geq \frac{8}{9},$$

which may be compared with $E = (\frac{2}{3})(l + \frac{1}{2})^2 [l(l + 1)] \geq 3 \frac{1}{2}$ when $k = \frac{1}{2}n$. The higher efficiency of $k = \frac{1}{2}n$ compared with $k = \frac{1}{3}n$ suggests the adoption of $k = \frac{1}{2}n$ in preference to $k = \frac{1}{3}n$ in general. Indeed its high efficiency in the case examined above indicates the occasional value of the simple method proposed even in cases where the least-squares method is available.

ASSESSMENT OF ACCURACY IN THE GENERAL CASE

In the general problem it is assumed that both y and x are subject to error. To use Wald's confidence interval method it is assumed further that the n errors $\eta \equiv y - Y$ are independently and normally distributed with constant variance σ_η^2 , similarly the n errors $\epsilon \equiv x - X$ are independent and normal with variance σ_ϵ^2 ; the x and y errors are moreover mutually independent, so that the variance of $\eta - \beta\epsilon$ is $\sigma_\eta^2 + \beta^2\sigma_\epsilon^2$.

Consider now possible 'estimates' of this last variance when β is known. If we write for the total sums of squares and products of x and y within the three groups

$$S_{xx} \equiv \sum_1 (x - \bar{x}_1)^2 + \sum_2 (x - \bar{x}_2)^2 + \sum_3 (x - \bar{x}_3)^2$$

$$S_{xy} \equiv \sum_1 (x - \bar{x}_1)(y - \bar{y}_1) + \sum_2 (x - \bar{x}_2)(y - \bar{y}_2) \\ + \sum_3 (x - \bar{x}_3)(y - \bar{y}_3)$$

$$S_{yy} \equiv \sum_1 (y - \bar{y}_1)^2 + \sum_2 (y - \bar{y}_2)^2 + \sum_3 (y - \bar{y}_3)^2,$$

where \sum_i denotes summation over the observations in the i -th group, then $(S_{yy} - 2\beta S_{xy} + \beta^2 S_{xx}) / (n - 3)$ is an estimate of the variance $\sigma_\eta^2 + \beta^2\sigma_\epsilon^2$ with $n - 3$ degrees of freedom. The remaining 3 degrees of freedom are contained in the three group means. One is represented by the general mean, one by the difference between the means of the first and third groups to be used in the estimate of β ; the third is represented by the difference between the mean of the second group and the general mean of the first and third groups.

For data with few observations it is advisable to make use of the last degree of freedom in the variance estimate, as in the numerical example considered later. Alternatively, if it is not so used, it remains available for testing the linearity of the true X, Y relation. In the former case,

the appropriate square to be added to the numerator of the previous estimate is

$$\{(\bar{y}_1 + \bar{y}_3 - 2\bar{y}_2)^2 - 2\beta(\bar{y}_1 + \bar{y}_3 - 2\bar{y}_2)(\bar{x}_1 + \bar{x}_3 - 2\bar{x}_2) + \beta^2(\bar{x}_1 + \bar{x}_3 - 2\bar{x}_2)^2\} \left\{ \frac{2}{k} + \frac{1}{n-k} \right\}$$

and the estimate $s^2(\beta)$ obtained with $n-2$ now as the divisor will have $n-2$ degrees of freedom.

Since

$$(\bar{x}_3 - \bar{x}_1)(b' - \beta) = (\bar{\eta}_3 - \beta\bar{\epsilon}_3) - (\bar{\eta}_1 - \beta\bar{\epsilon}_1),$$

when b' is given by (2), the left-hand quantity under the assumptions made in this section is normal with variance $(\sigma_\eta^2 + \beta^2\sigma_\epsilon^2)(2/k)$. This is subject to one qualification, that the errors in the x variable do not influence the allocation of the observations to the three groups. Such an effect may be neglected in many problems, particularly when the errors are small compared with the spacing of the observations at the points of division between the three groups; it will not be considered further here. A more detailed consideration of this point has been given by Wald (1).

Under the same assumptions we have

$$t = \frac{(\bar{x}_3 - \bar{x}_1)(b' - \beta)\sqrt{\frac{1}{2}k}}{s(\beta)}.$$

Although the denominator depends on β , this t -variate enables a confidence interval to be obtained for β . Thus for a value t corresponding to any chosen probability value we have the interval determined by the quadratic equation for β ,

$$(5) \quad (\bar{x}_3 - \bar{x}_1)^2(b' - \beta)^2\frac{1}{2}k = t^2(s_y^2 - 2\beta s_{xy} + \beta^2 s_x^2),$$

where $s^2(\beta) = s_y^2 - 2\beta s_{xy} + \beta^2 s_x^2$.

If required, a similar method may be used to provide a joint confidence region for α and β . If $a \equiv \bar{y} - \beta\bar{x}$, then a is independent of the numerator of t and of $s(\beta)$, and hence

$$F = \frac{\frac{1}{2}\{n(a - \alpha)^2 + \frac{1}{2}k(\bar{x}_3 - \bar{x}_1)^2(b' - \beta)^2\}}{s^2(\beta)}$$

is a variance ratio with degrees of freedom 2, $n-2$. For any chosen probability value the corresponding critical value of F will determine an ellipse as the boundary of the confidence region for α and β . This may be compared with the corresponding region for the least-squares method if it is known that $\sigma_\epsilon^2 = 0$; this region is similarly obtained from the variance ratio

$$F = \frac{\frac{1}{2}\{n(a - \alpha)^2 + (b - \beta)^2 \sum (x - \bar{x})^2\}}{s^2},$$

where s^2 is the usual variance estimate of $y - Y$ obtained from the residuals of y with $n - 2$ degrees of freedom.

If, as suggested earlier in this section, it is desired to examine the linearity of the functional relation, the variance estimate $s_{\beta}^2(3)$ of $\sigma_y^2 + \beta^2 \sigma_x^2$ with $n - 3$ degrees of freedom must be used. The further quantity

$$t = \frac{\{(\bar{y}_1 + \bar{y}_3 - 2\bar{y}_2) - \beta(\bar{x}_1 + \bar{x}_3 - 2\bar{x}_2)\} \left\{ \frac{2}{k} + \frac{1}{n-k} \right\}^{-\frac{1}{2}}}{s_{n-3}(\beta)}$$

is then (if the linear relation is valid) also a t -variate with $n - 3$ degrees of freedom. It will be noticed that it involves the unknown slope β . When this is replaced by the estimate b' , the resulting statistic is no longer exactly a t -variate, but might be treated approximately as such, especially when $\bar{x}_1 + \bar{x}_3 - 2\bar{x}_2$ is small compared with $\bar{x}_3 - \bar{x}_1$.

NUMERICAL EXAMPLE

As a numerical example consider fitting a straight line to the data on penicillin 'assay' given by Davies (3, § 6.12). Six different concentrations of pure penicillin were set up on a plate on which an agar medium containing *B. subtilis* had been spread, and the mean circle diameters of the zones of inhibition of growth of the organisms were measured (for further details of the technique see § 5.41 of (3)). The concentration had negligible error, so that the standard least-squares method was available, the relation between circle diameter and log. concentration being linear. With circle diameter y in mms. and 1 penicillin unit per ml. as $x = 1$, and a two-fold increase in concentration as the unit for the x scale, the regression equation of y on x was

$$(6) \quad Y = 20.403 + 1.782(x - 3.5) = 14.166 + 1.782x$$

with a 95% confidence interval for the slope, based on the usual t -statistic, of (1.732, 1.832).

It is stressed that the data are considered again here purely in order to illustrate the present method. The six observations are divided into three groups:

y	15.87	17.78,	19.52	21.35,	23.13	24.77	(Total 122.42)
x	1	2,	3	4,	5	6	(Total 21)

$$b' = \frac{(24.77 + 23.13) - (17.78 + 15.87)}{(6 + 5) - (2 + 1)} = 1.781.$$

Hence the estimated relation is

$$(7) \quad Y = 20.403 + 1.781(X - 3.5) = 14.170 + 1.781X.$$

The sum of squares within each group has only one degree of freedom in this example, and may conveniently be calculated from the difference of the two observations per group. The other degree of freedom to be added is that for the contrast of the mean for the second group with the mean for the other two groups. This gives zero contribution for x , and for y

$$24.77 + 23.13 + 15.87 + 17.78 - 2(19.52 + 21.35) = -0.19$$

with appropriate divisor. Hence

$$4s_x^2 = \frac{(1.91)^2 + (1.83)^2 + (1.64)^2}{2} + \frac{(-0.19)^2}{12} = 4.8463$$

$$4s_{xy} = \frac{1 \times 1.91 + 1 \times 1.83 + 1 \times 1.64}{2} + \frac{(0 \times -0.19)}{12} = 2.69$$

$$4s_y^2 = \frac{1^2 + 1^2 + 1^2}{2} + \frac{0^2}{12} = 1.5.$$

Equation (5), with $t = 2.78$ for 4 degrees of freedom ($P = 0.05$), gives

$$16(1.781 - \beta)^2 = (2.78)^2(4.8463 - 2\beta \times 2.69 + 1.5^2)/4$$

$$\text{or} \quad 13.1018\beta^2 - 2\beta(23.2987) + 41.3879 = 0$$

$$\text{or} \quad \beta = 1.778 \pm 0.058.$$

Thus the 95% confidence interval for β by this method is (1.720, 1.836), an interval naturally slightly wider than the interval obtained by the least-squares method, since the assumption of no error in x has been dropped.

REFERENCES

- (1) Wald, A. The Fitting of Straight Lines if Both Variables are Subject to Error. *Ann Math Stat.* 11, 284, 1940.
- (2) Lindley, D. V. Regression Lines and the Linear Functional Relationship. *J. Roy. Stat. Soc. (Suppl.)* 9, 218, 1947.
- (3) Davies, O. L. (Editor). *Statistical Methods in Research and Production*. Oliver and Boyd, 1947.
- (4) Nair, K. R. and Shrivastava, M. P. On a Simple Method of Curve Fitting. *Sankhyā* 6, 121, 1942.
- (5) Nair, K. R. and Banerjee, K. S. A Note on Fitting of Straight Lines if Both Variables are Subject to Error. *Sankhyā* 6, 331, 1942.

RELATIONSHIP OF CATCH TO CHANGES IN POPULATION SIZE OF NEW ENGLAND HADDOCK

BY HOWARD A. SCHUCK

Aquatic Biologist

*Fish and Wildlife Service
United States Department of the Interior*

INTRODUCTION

THE UNITED STATES catch of haddock has fluctuated considerably throughout the years and these fluctuations have generally been of a declining nature. In 1929, the catch was about 260 million pounds, and in recent years it has averaged only about 150 million pounds. Fluctuations in the catch have been due in large part to variations in actual abundance, or the size of the stock of commercial sizes of haddock on the banks in different years. We are, therefore, interested in obtaining an accurate measure of the size and the composition of the stock, to measure its changes throughout the years, and to determine what factors have been most responsible for such changes. Changes in the stock from year to year are the result of varying rates of removals and additions. Therefore, besides determining the size of the stock in different years, it is necessary to measure the yearly removals from the stock by catch and natural mortality, and the yearly additions by recruitment and growth.

If these variables could be measured accurately, we should be in a position to evaluate their relative importance in determining the size of the stock and to determine whether the size of the spawning stock and of other stocks affects recruitment. With such information and other general life history facts, it should be possible to determine at what level the stock should be maintained, to determine what mode and intensity of fishing will result in the maximum sustained production of haddock, and to make periodic predictions as to future production of the fishery for the industry.

A basic equation is:

$$S + (G + R + M) - (C + N + M_1) = S_1$$

where:

S = size of population at the beginning of the year.

S_1 = size of population at the end of the year.

G = additions to the population during the year by growth.

R = additions to the population during the year by recruitment of young.

M = additions due to immigrations.

C = deductions from the population during the year by the fishery.

N = deductions from the population during the year due to natural mortality.

M_1 = deductions due to emigrations.

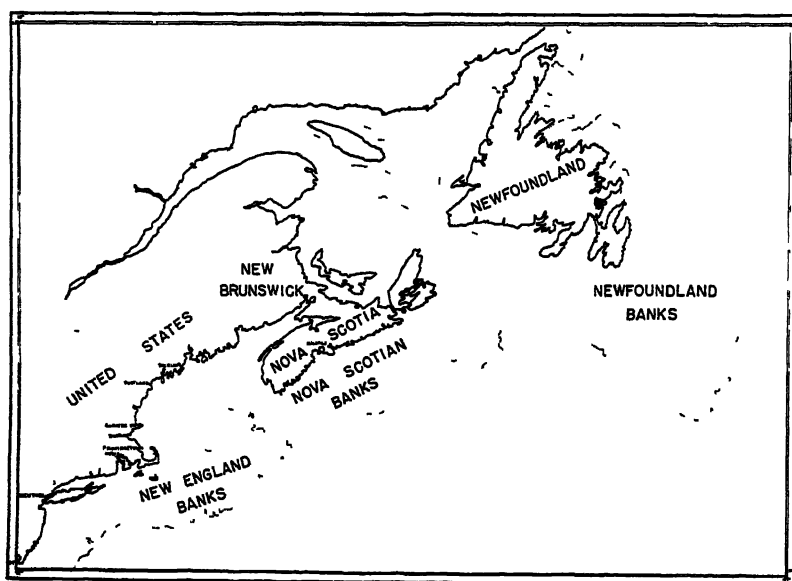


FIGURE 1.
LOCATION OF FISHING BANKS OFF NEW ENGLAND, NOVA SCOTIA, AND
NEWFOUNDLAND.

It is believed that the population of haddock inhabiting the New England Banks (Georges) (Fig. 1) is largely independent of the populations on the Nova Scotian and Newfoundland Banks. Assuming this to be true, and if we consider the population on Georges Bank only, there will be no important changes in the stock from year to year due to

TABLE 1
RELATIVE SIZE OF THE GEORGES BANK HADDOCK POPULATIONS IN TERMS OF
THE AVERAGE NUMBERS OF FISH PER DAY TAKEN BY A STANDARD GROUP OF
OTTER TRAWLERS

Year	Numbers per day
1931	3,032
1932	4,324
1933	3,630
1934	4,049
1935	4,927
1936	5,590
1937	4,404
1938	4,833
1939	5,502
1940	4,979
1941	6,960
1942	7,941
1943	7,319
1944	5,737
1945	5,347
1946	4,956
1947	4,954
Average	5,205

immigrations or emigrations; and M and M_1 can be left out of the equation.

Also, if we consider the population as numbers, rather than pounds of fish, G or "growth", can be left out too. Furthermore, if we define the population S as being the number of fish of certain year classes at the beginning of a year and S_1 as the number of fish of the same year classes at the end of that year, then there can be no recruitment; and R can also be ignored. Thus, the equation for certain purposes can be reduced to:

$$S - (C + N) = S_1$$

Available for use in this equation are biological and statistical data for the Georges Bank population going back to 1931. These data were assembled by the Haddock Investigation of the United States Fish and Wildlife Service and its predecessor agency, the United States Bureau of Fisheries.

The remainder of this paper will be devoted to: (1) developing an index representing the size of the population in terms of numbers of

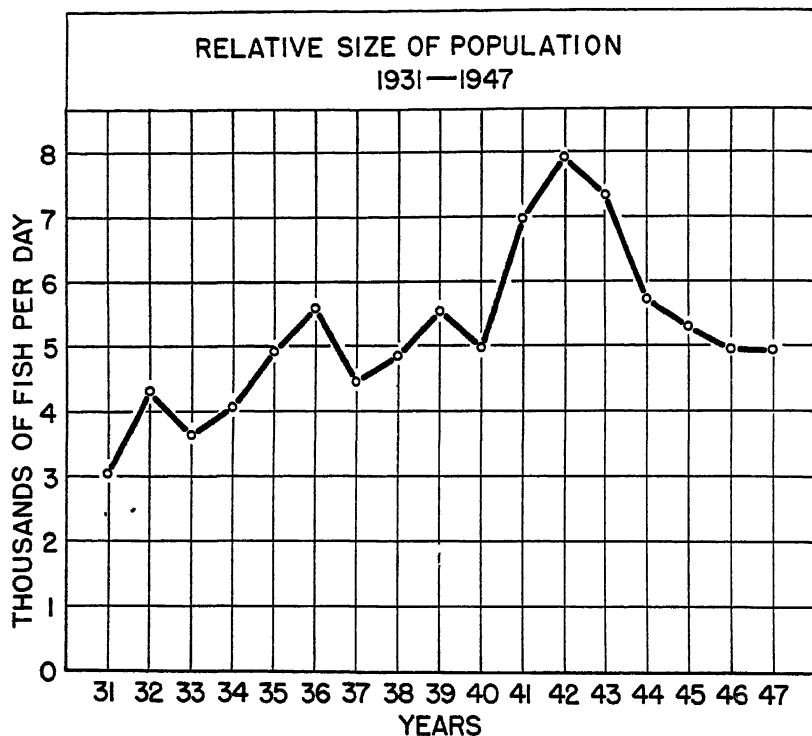


FIGURE 2.
RELATIVE SIZE OF THE POPULATION, IN TERMS OF THOUSANDS OF FISH PER DAY
BY YEARS.

haddock of definite ages and year classes, at the beginning and end of yearly periods (S and S_1); (2) measuring the fishery removals (C) of haddock of each age during each of the 17 years, 1931-47; and (3) determining how important the yearly fishery removals are in decreasing the stock from the beginning to the end of yearly periods.

SIZE OF THE STOCK OR " S " AND " S_1 "

Total catch represents fishing removals and in itself is a vital piece of information. It does not, however, represent abundance, or the relative size of the population on the Bank, inasmuch as the amount of fishing effort utilized to make the catch varies among years.

The index representing the relative size of the population that was used in the Haddock Investigation is the average yearly catch per day¹ of a standard group of large otter trawlers which fished out of Boston

¹Details of this abundance analysis were developed by W. C. Herrington and G. A. Rounsefell.

during this 17-year period. The relative size of the population² was first expressed in terms of the average number of pounds per day taken by these trawlers in each year. By the use of yearly average weight data, the statistics on relative population size were converted from pounds to numbers of fish (Table 1 Fig. 2).

In each year and season a *sample* of the haddock that were landed had been obtained, and from those fish obtained, scales had been collected. Then, for each year and season, the ages of these sample fish were determined. This determination was made by studying the projected impression of these scales. Figure 3 shows a photograph of such a microprojection of a scale from a Georges Bank haddock.

The fish were aged as having completed their first, second, third, fourth, fifth, sixth, seventh, eighth, and ninth year of life, and were correspondingly classified as fish of 1 to 9 years of age. The category of 9-year-olds includes 9-year-old and older fish. (The number of haddock of ages greater than 9 years was very small, amounting in the aggregate to less than one-half of one percent of all haddock in the catch.)

By using the percentage age composition that had been computed for haddock of each length and for each year and season, the total numbers of fish caught per day were reduced to numbers per day of each age (Table 2 and Fig. 4). The average abundance for all 17 years (Fig. 5) amounted to 116 one-year-olds, 1,472 two-year-olds, 1,571 three-year-olds, 920 four-year-olds, 557 five-year-olds, 324 six-year-olds, 149 seven-year-olds, 61 eight-year-olds and 34 nine-year-old and older fish. It can be concluded that the relative abundance of fish in the catch diminishes quite regularly for those fish three years old and older. The fact that the one- and two-year-old fish are less abundant indicates that these age groups are not fully available to the fishery.

In order to measure the diminution in the stock over the period of a year, it was desired to compare the relative population size of the fully-available age groups of fish at the beginning of each year with the size of the corresponding stock at the end of the year. Table 2 gives the average population size for the "haddock" year.³ In order to obtain a

²Where "population size" is mentioned in the remainder of this paper it refers to this index of relative population size. Although it has not yet been possible to determine the exact relationship between the actual number or pounds of fish in the stock and our calculated index of relative population size, the index appears to suffice for the purpose used here.

³The "haddock" year consists of seasons A, B, C, and D as follows:

A—February, March, and April (spawning season)

B—May, June, and July

C—August, September, and October

D—November, December, and January



FIGURE 3

PHOTOGRAPH OF A SCALE FROM A HADDOCK THAT HAD JUST COMPLETED ITS FOURTH YEAR WHEN CAUGHT APRIL 1939 ON GEORGES BANK. THE MARKS INDICATE THE COMPLETION OF EACH YEAR OF GROWTH. THE LENGTH OF THIS FISH WAS 22 3/4 INCHES

TABLE 2
RELATIVE POPULATION OF EACH AGE OF GEORGES BANK HADDOCK IN TERMS OF
NUMBERS CAUGHT PER DAY

Year	Age in years									
	All Ages	1	2	3	4	5	6	7	8	9 and older
1931	3,032	147	691	158	439	699	466	256	132	44
1932	4,324	11	210	2,829	275	413	323	146	74	43
1933	3,630	44	986	720	1,145	249	193	145	67	81
1934	4,049	141	960	1,106	690	678	241	125	60	40
1935	4,927	202	1,704	1,306	574	509	428	97	74	33
1936	5,590	157	1,752	1,834	920	402	236	222	41	26
1937	4,404	150	1,233	1,327	698	535	251	119	65	26
1938	4,533	165	2,590	988	489	234	199	114	31	23
1939	5,502	95	1,775	2,416	640	271	123	108	42	32
1940	4,970	324	1,116	1,089	1,018	309	184	93	28	18
1941	6,900	144	3,298	1,275	1,046	752	233	123	40	49
1942	7,941	94	3,036	2,567	1,037	624	362	158	36	27
1943	7,319	11	1,026	3,470	1,551	530	492	149	61	29
1944	5,737	14	135	1,412	2,609	948	416	95	91	17
1945	5,347	25	1,663	420	1,244	1,218	485	194	01	37
1946	4,956	24	856	1,992	400	854	562	217	49	2
1947	4,954	18	1,996	1,180	863	250	314	180	89	55
Total	88,484	1,966	25,033	26,700	15,638	9,473	5,508	2,541	1,041	582
Avg.	5,205	116	1,472	1,571	920	557	324	149	61	34

value of the population size at the beginning of the year while eliminating the effect of the seasonal cycle in availability it was necessary to recompute these data.

All data originally had been computed on a seasonal basis: for example, Table 3 shows the seasonal population size data from which Table 2 was derived. In order to obtain values for the population size that more closely represent values at the beginning of each year,⁴ the abundance values for seasons C, D, A, and B in Table 3 were averaged. In this recombination it was necessary to consider that 3-year-old fish in seasons C and D become 4-year-old fish in seasons A and B of the following year and that other ages progress accordingly.

For example, to obtain the relative size of the population of 4-year-old fish at the beginning of 1935 the following figures⁵ were used:

⁴It is recognized that by summarizing values for 4 seasons and dividing by 4, the average obtained does not under some conditions represent the average of the midpoint and thus the exact beginning of the year. For the purpose of this analysis, however, such a calculation represents the beginning of the year accurately enough.

⁵An exception to this rule was made in computing the relative population of 9-year-old haddock at the beginning of the year. Since this group includes all older haddock, 8-year-old and 9-year-old haddock from seasons C and D were added to 9-year-old haddock from seasons A and B and the total of these 6 figures, instead of the usual 4, was divided by 4 to give the average.

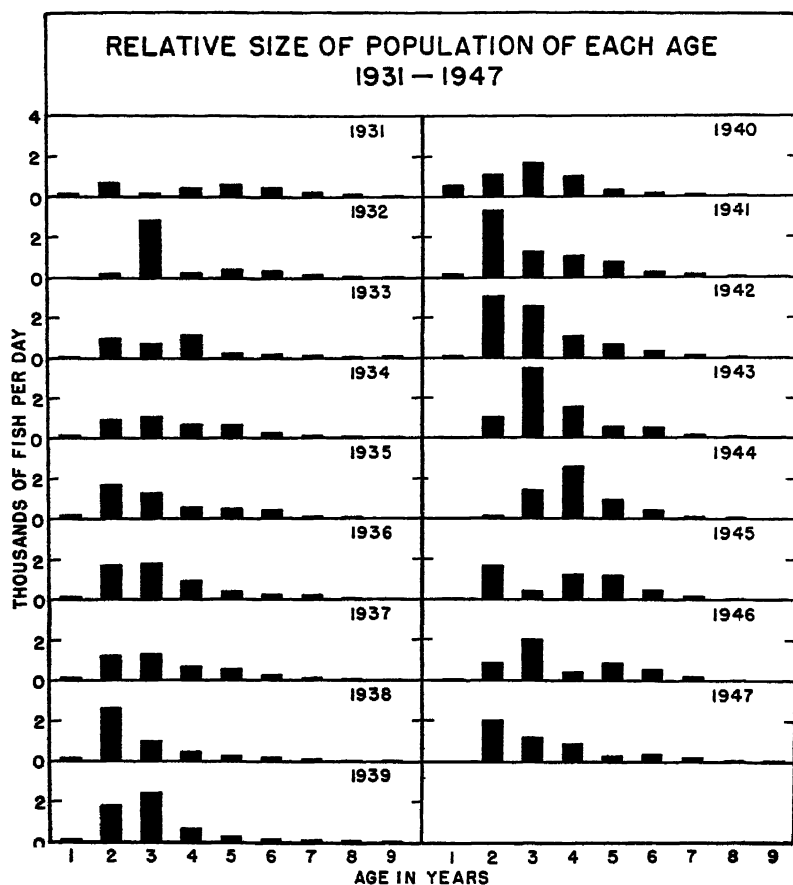


FIGURE 4.

RELATIVE SIZE OF HADDOCK POPULATION OF AGES 1-9, FOR EACH OF THE 17 YEARS.

	Number of fish per day
3-year-old haddock, season <i>C</i> , 1934	1,425
3-year-old haddock, season <i>D</i> , 1934	431
4-year-old haddock, season <i>A</i> , 1935	583
4-year-old haddock, season <i>B</i> , 1935	889
Total	3,328
Average	832

Using this system the relative sizes of the population of 4- to 9-year-old fish at the beginning of each year were computed and are shown in

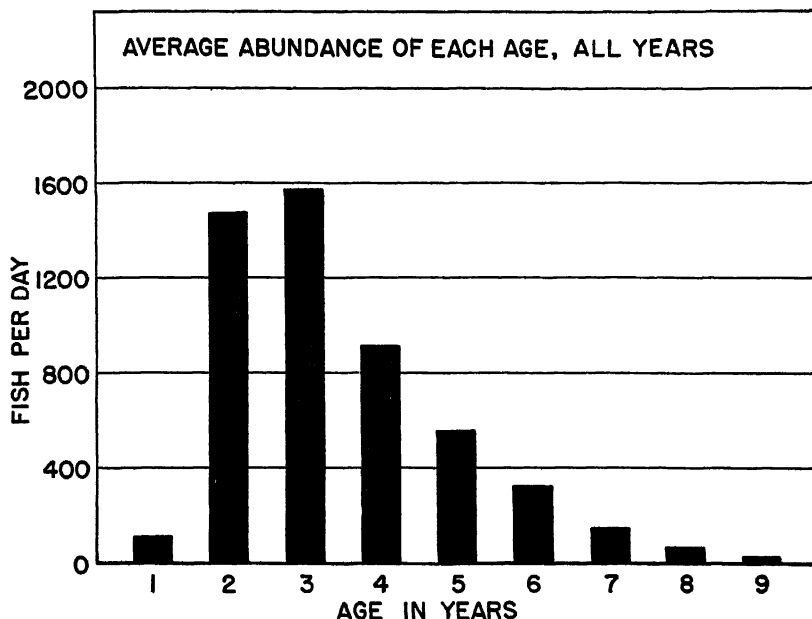


FIGURE 5.

RELATIVE POPULATION SIZE OF HADDOCK OF AGES 1-9 AVERAGE OF ALL 17 YEARS

Table 4. Since computation of the catch-per-day of 3-year-old fish at the beginning of the year involved use of figures for the less available 2-year-old fish in seasons *C* and *D*, it was decided to omit the 3-year group.

The next step was to decide whether to consider the age groups separately or in the aggregate. Examination of the data in Table 3 indicated that the decrease in catch-per-day for individual year classes from year to year was rather variable, hence, it was desirable to combine age groups. Thus the total of all fish of ages 4 to 9 years for the beginning of each year are shown in the right-hand column of Table 4.

It was next necessary to compute the size of the population at the end, in addition to at the beginning, of each year. The average population at the beginning of the year or seasons $(C + D + A + B)/4$, approximates the value of the midpoint between *D* and *A*. Therefore, values for the number of fish at the beginning of the year are the same as values for the number at the end of the preceding year.

For example, from Table 4, if there are 1,793 five-year-old fish per day at the beginning of 1945, then there are 914 (the number of 6-year-olds at the beginning of 1946) 5-year-olds at the end of 1945.

TABLE 3
RELATIVE SIZE OF POPULATION OF EACH AGE OF GEORGES BANK HADDOCK
BY SEASONS IN NUMBERS CAUGHT PER DAY

Year	Season	Age in years									
		No. all years	1	2	3	4	5	6	7	8	9 and older
1931	A	3,268	.	30	103	560	1,088	781	372	155	89
	B	3,182	.	51	70	770	1,139	630	336	110	37
	C	2,562	.	897	146	245	348	333	269	224	40
	D	3,114	587	1,755	183	179	222	99	45	32	12
1932	A	4,281	.	11	2,418	489	707	387	143	80	46
	B	4,937	.	38	3,333	359	410	351	275	95	76
	C	5,848	3	430	4,253	149	343	458	96	91	25
	D	2,231	41	301	1,310	103	191	96	70	32	27
1933	A	3,697	.	112	254	1,728	411	423	324	183	262
	B	4,349	.	1,318	494	1,814	277	185	161	56	44
	C	4,487	39	1,724	1,461	875	207	93	63	17	8
	D	1,985	138	789	671	164	100	72	31	12	11
1934	A	3,729	.	4	1,360	471	874	574	198	150	98
	B	4,299	.	290	1,217	1,368	866	209	252	41	56
	C	4,619	.	1,929	1,425	544	302	148	33	37	1
	D	3,549	565	1,640	431	377	472	33	17	10	4
1935	A	3,215	.	23	884	553	748	607	185	86	90
	B	5,536	.	1,117	1,719	889	740	884	52	131	4
	C	5,495	16	2,443	1,848	590	367	152	67	9	3
	D	5,462	791	3,235	773	234	179	71	84	71	24
1936	A	5,827	.	267	2,078	1,688	903	397	356	47	91
	B	7,217	4	1,760	3,118	997	517	341	377	103	.
	C	6,171	93	3,615	1,537	692	41	69	110	2	9
	D	3,143	532	1,381	603	303	145	138	45	11	5
1937	A	5,224	1	423	1,810	1,167	988	500	176	137	22
	B	4,969	.	1,068	1,838	949	552	282	161	54	45
	C	5,175	361	2,757	1,265	439	243	24	51	21	14
	D	2,247	237	655	375	237	356	196	90	46	25
1938	A	3,078	.	363	1,015	694	341	390	159	56	60
	B	4,736	.	2,133	1,233	614	288	270	178	13	7
	C	7,425	.	5,500	1,241	384	162	54	52	26	6
	D	4,092	660	2,363	463	264	144	81	68	28	21
1939	A	4,463	.	257	2,260	873	515	191	179	66	92
	B	6,629	.	1,604	3,371	953	251	223	143	66	13
	C	6,848	30	3,057	3,157	424	148	16	13	3	.
	D	4,066	349	2,150	877	307	174	56	96	34	23

Also, the aggregate population of 4-year-old and older fish at the beginning of a year would produce survivors at the end of the year which would amount to the number of 5-year and older fish at the beginning of the next year. For example, if the total population of 4-year-old and older fish at the beginning of 1944 was the total of 3,188 four-year-

TABLE 3—Continued

Year	Season	Ages in years									
		No. all years	1	2	3	4	5	6	7	8	9 and older
1940	A	2,805	..	127	1,057	1,046	337	141	52	24	21
	B	6,245		1,200	2,449	1,521	509	234	268	42	22
	C	5,638	221	2,175	1,832	922	161	289	10	23	5
	D	5,228	1,876	961	1,415	582	230	73	43	24	24
1941	A	5,855	1	1,463	1,735	1,289	1,003	181	135	22	26
	B	7,692	.	2,165	1,732	1,639	1,222	497	209	96	132
	C	9,082	128	6,498	1,075	757	381	123	81	21	18
	D	5,210	448	3,068	557	498	403	131	66	19	20
1942	A	5,863	...	290	2,599	1,083	1,006	598	224	35	28
	B	8,769	.	1,636	3,780	1,674	858	444	244	74	59
	C	9,058	39	4,875	2,898	703	212	232	93	3	3
	D	8,074	336	5,344	989	688	421	173	73	31	19
1943	A	7,007	...	116	3,282	2,069	688	668	102	114	28
	B	8,247	..	380	4,048	1,844	831	689	287	97	71
	C	7,316	...	1,892	3,353	1,277	429	291	59	7	8
	D	6,647	45	1,716	3,194	1,013	173	320	146	29	11
1944	A	7,107	..	1	1,207	3,622	1,221	680	158	202	16
	B	6,029	...	35	1,396	2,586	1,385	448	64	81	34
	C	6,792	..	309	2,362	3,055	730	224	78	34	...
	D	3,020	54	193	681	1,174	458	312	80	48	20
1945	A	4,687	.	33	499	1,023	1,851	762	351	138	30
	B	5,640	..	1,555	374	1,862	1,007	469	193	12	78
	C	7,293	34	3,277	668	1,502	1,133	462	162	34	21
	D	3,769	67	1,783	138	589	792	248	69	58	25
1946	A	4,768	...	45	1,940	487	1,238	804	140	113	1
	B	6,030	..	735	3,040	333	1,059	928	446	80	9
	C	4,403	45	1,306	1,723	310	335	268	215	...	1
	D	4,024	52	1,334	1,270	471	582	246	60
1947	A	4,382	...	58	1,253	1,028	394	507	293	149	100
	B	3,569	...	1,052	1,015	600	264	337	175	94	52
	C	8,122	1	4,904	1,831	782	169	253	115	41	26
	D	3,721	73	1,960	658	442	171	159	136	71	42
Avg. all years	A	4,666	...	215	1,521	1,206	842	505	209	103	65
	B	5,806	...	1,069	2,015	1,223	721	436	225	74	43
	C	6,255	59	2,790	1,880	805	359	206	92	35	11
	D	4,093	403	1,807	858	448	307	147	72	33	18

olds, 1,224 five-year-olds, 432 six-year-olds, 208 seven-year-olds, 122 eight-year-olds, and 26 nine-year-old and older fish, or a total of 5,200; then the survivors from this group of year classes, after an interval of one year, would be the number of 5- to 9-year-old and older fish at the beginning of 1945, or 1,793 five-year-olds, 604 six-year-olds, 270 seven-

TABLE 4
SIZE OF HADDOCK POPULATION AT BEGINNING OF EACH YEAR¹

Year	Age in years						Total
	4	5	6	7	8	9 and older	
1932	304	385	327	218	122	108	1,464
1933	2,276	235	286	260	101	120	3,278
1934	993	695	272	154	71	51	2,236
1935	832	602	616	104	67	39	2,260
1936	1,326	561	321	239	75	50	2,572
1937	1,064	634	242	136	86	24	2,186
1938	737	326	315	139	52	43	1,612
1939	884	354	180	114	63	46	1,641
1940	1,650	394	174	98	44	26	2,386
1941	1,544	932	267	176	43	58	3,020
1942	1,097	780	456	181	64	42	2,620
1943	1,950	728	498	198	94	39	3,507
1944	3,188	1,224	432	208	122	26	5,200
1945	1,482	1,793	604	270	77	52	4,278
1946	406	1,097	914	324	106	37	2,884
1947	1,305	360	490	246	132	38	2,571
Total	21,038	11,100	6,394	3,065	1,319	799	43,715
Average	1,315	694	400	192	82	50	2,733

¹Values are the average of the number of fish of the particular age from Seasons A and B of the year in question, and of the number of fish of 1 year younger from Seasons C and D of the preceding year.

year-olds, 77 eight-year-olds, and 52 nine-year-old and older fish, or a total of 4,278.

Having already obtained the total number of 4-year-old and older fish at the beginning of the year (from Table 4), and having now computed the total number of 4-year-old and older fish at the end of each year (the number of 5-year-old and older at the beginning of the next year), all such totals were entered in Table 5.

Computation of the yearly diminution of the stocks being measured was then only a matter of subtracting the value representing the stock at the end of the year from the value representing the stock at the beginning of each year.

THE FISHERY REMOVALS OR "C"

A measure of the yearly decreases in population size of completely available fish from year to year thus had been obtained. Inasmuch as

TABLE 5
RELATIVE SIZE OF POPULATION OF CERTAIN AGES OF HADDOCK AT THE
BEGINNING AND END OF EACH 15 YEARS

Year	Number of 4- to 9-year-olds at beginning of year	Number of 4- to 9-year-olds at end of year ¹	Decrease
1932	1,464	1,002	462
1933	3,278	1,243	2,035
1934	2,236	1,428	808
1935	2,260	1,246	1,014
1936	2,572	1,122	1,450
1937	2,186	875	1,311
1938	1,612	757	855
1939	1,641	736	905
1940	2,386	1,476	910
1941	3,020	1,525	1,495
1942	2,620	1,557	1,063
1943	3,507	2,012	1,495
1944	5,200	2,796	2,404
1945	4,278	2,478	1,800
1946	2,884	1,266	1,618
Total	41,144	21,519	19,625
Average	2,743	1,435	1,308

¹End of year = number of 5- to 9-year-olds at beginning of following year.

the purpose of this study was to determine to what extent such decreases were associated with, or were the result of, the removals by the fishery, it was necessary next to determine how many fish the fishery had taken from the population in the various years.

The fishery removals for the years 1931-47⁶ were first tabulated in terms of pounds of fish. Having also the average weights of these fish that were landed, the total numbers caught were easily computed. The total pounds and numbers are shown in Table 6, and the total numbers in Figure 6.

The numbers caught were then reduced to numbers of each age by utilizing the percentage-age compositions referred to earlier. After summarizing by size groups and season, the number of fish of each age removed by the fishery in each of the 17 years is shown in Table 7.

⁶The landings for the ports of Boston, Gloucester, New Bedford, Mass., and Portland, Me.

TABLE 6
TOTAL CATCH OF HADDOCK FROM NEW ENGLAND BANKS

Year	Millions of pounds	Millions of fish
1931	101.801	34.979
1932	86 706	32.348
1933	70 272	26 623
1934	39 683	15.617
1935	68 579	28.565
1936	73.496	31.489
1937	83 973	32 528
1938	80 202	33.570
1939	91.181	38.911
1940	81 676	31.345
1941	111 611	46.944
1942	97 786	41.299
1943	80 215	33.036
1944	84 265	29.062
1945	65.284	22.091
1946	90 802	32.678
1947	98 082	38.931
Total	1,405.614	550.016
Average	82 683	32 354

DECLINE IN THE SIZE OF STOCK AS ASSOCIATED WITH
VARIATIONS IN THE CATCH.

In an earlier section of this paper the yearly declines in the relative size of the stocks of those ages of haddock that were fully available to the fishery were computed (Table 5). In the section just completed, the catch of fish of each age in each year has been computed (summarized in Table 7). By summing the catches of fish of 4 to 9 years of age inclusive, in each year, the numbers of fish that were removed from the corresponding stock between the beginning and the end of the years involved were computed. Thus, in Table 8 data are presented which represent:

- (1) the decrease in the relative size of the stocks of 4- to 9-year-old fish from the beginning to the end of each of the 15 years (1932-46) in thousands of fish per day, and
- (2) the number of fish removed by the fishery from these stocks during each of these 15 yearly intervals, in millions.

REMOVALS FROM THE GEORGES BANK POPULATION 1931 — 1947

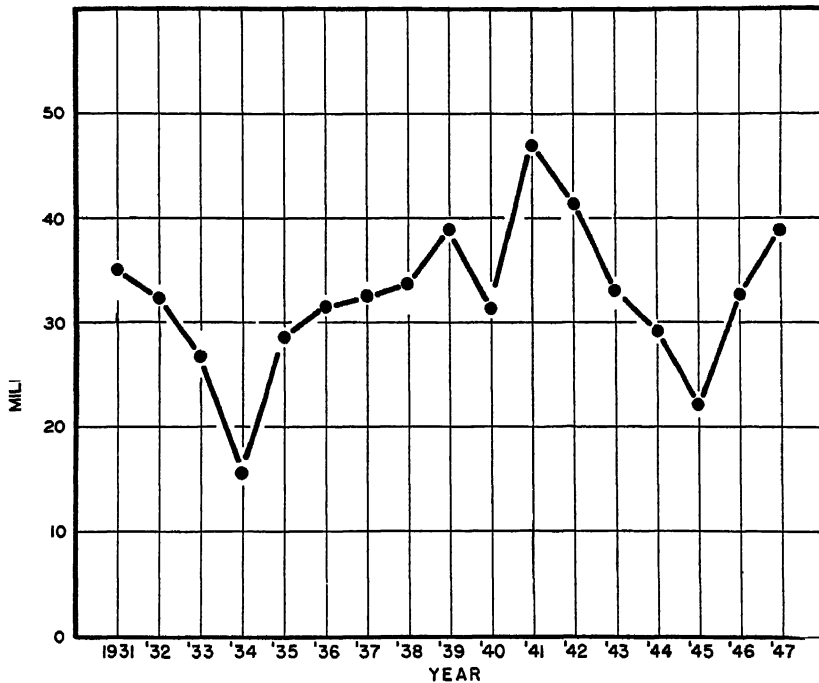


FIGURE 6.

CATCH OF GEORGES BANK HADDOCK IN TERMS OF NUMBERS OF FISH.

A casual observation of this table shows, in general, that in years during which large numbers of fish were taken from the Georges Bank population, there were also large declines in the population size from the beginning to the end of the year, and that in years during which small numbers were removed, the population changed but little.

These data have been plotted in Figure 7, with the "removal" or "catch (C)" as the independent variable, and with the decline, the change in population size from the beginning to the end of the year, as the dependent variable. This figure is plotted in a rather unusual manner, with values of the dependent variable being plotted below, rather than above the origin. This has been done inasmuch as values of the dependent variable (change in population size) are actually decreases rather than increases, and it has been found that this method of plotting is more easily interpreted by some people.

TABLE 7
AGE COMPOSITION OF CATCH, BY YEARS, IN MILLIONS OF FISH

Year	Age in years									Total
	1	2	3	4	5	6	7	8	9 and older	
1931	1 061	8 167	1 802	5 089	7 075	5 291	2 949	1 555	490	34 979
1932	099	1 712	21 139	2 006	3 031	2 383	1 079	553	324	32 348
1933	210	7 366	5 218	8 648	1 791	1 346	1 030	464	550	26 623
1934	296	3 807	4 470	2 889	2 518	525	482	197	133	15 617
1935	1 144	11 096	7 803	3 136	2 467	2 053	415	360	089	28 565
1936	528	11 449	10 171	4 629	1 803	1 153	1 140	217	099	31 459
1937	1 193	10 129	9 715	4 890	3 374	1 608	815	416	188	32 358
1938	961	18 453	6 966	3 304	1 568	1 312	765	198	143	33 570
1939	565	12 806	17 379	4 353	1 807	904	695	272	200	38 911
1940	1 805	6 692	11 061	7 261	2 188	1 286	653	191	117	31 345
1941	697	21 404	9 026	7 359	5 303	1 632	860	280	353	46 494
1942	290	13 106	14 577	5 938	3 648	2 131	941	205	162	41 299
1943	016	3 653	15 659	7 423	2 742	2 385	688	313	157	33 036
1944	054	675	7 410	13 047	4 945	1 991	435	412	093	29 062
1945	101	7 046	1 698	5 285	4 862	1 941	766	223	169	22 001
1946	191	6 709	13 251	2 406	5 000	3 289	1 589	224	019	32 678
1947	085	15 547	9 604	6 690	1 979	2 542	1 397	690	424	38 931
Total	10 259	159 817	167 149	94 388	57 221	33 972	16 700	6 770	3 710	550 016
Average	605	9 402	9 333	5 532	3 360	1 998	962	398	215	32 354

The straight line in Figure 7 was fitted to the data by the method of least squares and has the equation,

$$Y = -.022 + .1135X \text{ where}$$

X = millions of haddock of ages 4-9 years removed from the stock in each of 15 years by the fishery.

Y = decrease in relative population size of 4- to 9-year-old haddock during each of these 15 years in thousands of fish per day.

The coefficient of correlation measuring the degree of association between these two variables is 0.81. With 13 degrees of freedom this value proves to be highly significant (1 per cent level = 0.64). R^2 is about 0.66. Thus, it seems valid to conclude (under the assumption that the straight line best fits these data) that about 66 per cent of the variability in yearly decreases in population size, from the beginning to the end of the individual years, is explainable by the variations in the numbers of fish actually removed from the stock by the fishery. This value of 66 per cent is possibly a minimum estimate of the effect of the fishery, inasmuch as the index of abundance is probably not perfectly correlated with actual abundance.

No attempt was made in this treatment to determine whether some curved line fitted these data better than this straight line.

TABLE 8
DECREASE IN SIZE OF STOCK OF 4- TO 9-YEAR-OLD HADDOCK FROM
THE BEGINNING TO THE END OF YEARS 1932-46, AND THE TOTAL
CATCH OF THESE AGES IN EACH YEAR

Year	Decrease in stock thousands of fish per day	Catch in millions of fish
1932	.462	9 398
1933	2.035	13.829
1934	.808	7.044
1935	1.014	8.522
1936	1.450	9.041
1937	1.311	11.491
1938	.855	7.290
1939	.905	8.161
1940	.910	11.697
1941	1.495	15.817
1942	1.063	13.026
1943	1.495	13.708
1944	2.404	20 923
1945	1.800	13.246
1946	1.618	12.527
Total	19.625	175.720
Average	1.308	11.715

This line was arbitrarily extrapolated beyond the limits of the data towards the origin, although admittedly the exact position of the line where the removals are very small is unknown. It can be seen from Figure 7 that the intercept is practically at the 0.0 point. The position of this intercept, assuming that the population index actually represents the size of the population, poses interesting theoretical possibilities.

First of all, the suggestion is raised that within the ranges of population size and fishing removals represented by these data, the losses due to factors other than the fishing removals, i.e., to natural mortality, may be negligible. Such a possibility could theoretically be true under the conditions of an intensive fishery, where fishing removals would take many fish which would otherwise be removed by natural causes. When one takes into consideration the relative lack of bottom dwelling predators on Georges Bank that are large enough to consume haddock of 2-10 pounds and the apparent lack of any disease epidemic or serious parasitism in haddock over the 17-year period, this possibility does not appear quite so improbable.

DECREASE IN POPULATION SIZE AS AFFECTED BY CATCH

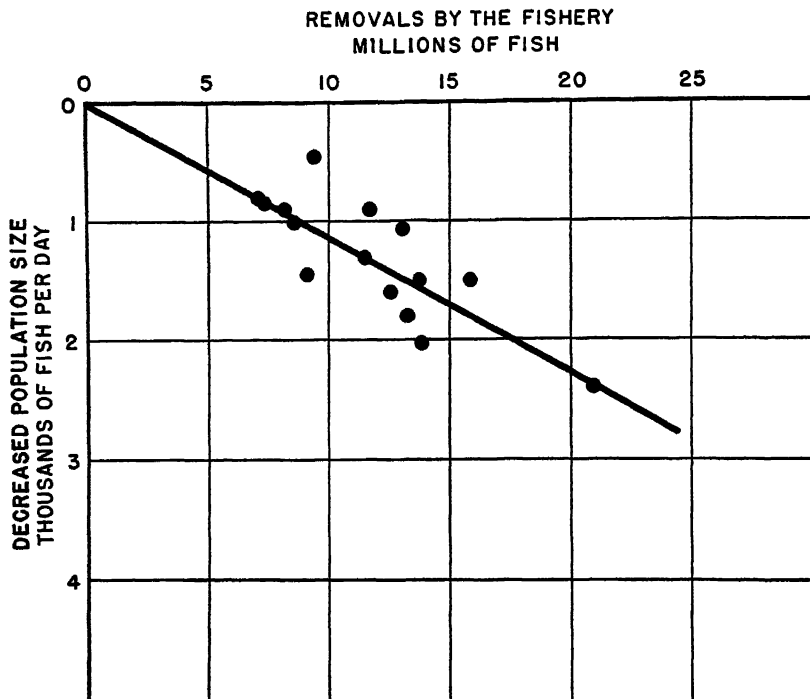


FIGURE 7.

THE RELATIONSHIP BETWEEN THE YEARLY REMOVALS IN MILLIONS OF HADDOCK AND THE DECREASE IN RELATIVE POPULATION SIZE FROM BEGINNING TO THE END OF THESE SAME YEARLY PERIODS IN TERMS OF THOUSANDS OF FISH PER DAY.

Secondly, the exact position of the line with populations of this general size, if the removals were greatly reduced and even became zero, is unknown. If the extrapolation, as in Figure 7, happens to be an accurate representation of this relationship, then one would conclude that with no fishing removals, as would occur if fishing were to suddenly cease, there would be no decrease in the stock and thus no natural mortality. Theoretically, however, if fishing were to be considerably reduced suddenly, natural mortality would probably be greater than at present because some of the fish now being caught would be vulnerable to whatever causes of natural mortality are in operation. With populations of present levels but with very small fishing removals, the line would

possibly curve toward the Y axis and intersect it at some point greater than 0.

The data and the ideas expressed here refer only to a heavily fished population and not to the relatively unfished populations of early days, or to the populations which would result if the sudden cessation of fishing were to continue for several years. In such populations, natural mortality would probably be greater yet, for such reasons as poorer nutrition of the larger stock, greater average age resulting in more deaths from senility, and so on.

This general situation is to be studied by various lines of approach in future studies. From the present study, however, we may conclude that the number of haddock caught in various years by the New England fleet markedly affected the subsequent population of haddock of corresponding ages on Georges Bank. Although it is generally assumed in many fisheries that the fishery does affect the stock, instances where such an effect has been demonstrated clearly are extremely rare. This analysis, in addition to demonstrating this relationship, is also of considerable value in providing the basic data that can be used in determining many other very important facts necessary for a broad understanding of the biometrics of the valuable New England haddock resource. Such facts include the actual number of fish present on the bank, fishing and natural mortality rates, growth rates, indices of the recruitment of young, the effect of various factors upon recruitment, and predictions as to the future abundance of this species. Investigations of these relationships are now being undertaken and will be reported upon soon.

ACKNOWLEDGMENTS

The system of collection of the basic biological and statistical data was developed by William C. Herrington, who until 1947 directed the North Atlantic Haddock Investigation. Employees of the United States Fish and Wildlife Service who aided the writer in summarizing the extensive data upon which this analysis is based include: James J. Miggins, Edgar L. Arnold, Jr., Louis D. Stringer, Frank A. Dreyer, Dorothy B. Monahan, Edward S. Phillips, Sterling L. Cogswell, Elizabeth V. Nugent and Margaret A. Reeves. Scales were read by Edgar L. Arnold, Jr., and the writer. Interpretation of certain data collected in early years was facilitated through the assistance of John R. Webster. The manuscript was reviewed critically by Dr. William F. Royce and Dr. Ralph P. Silliman of the Branch of Fishery Biology, United States Fish and Wildlife Service.

ONE DEGREE OF FREEDOM FOR NON-ADDITIVITY*†

JOHN W. TUKEY

Princeton University

INTRODUCTION

IN DISCUSSING the possible shortcomings of the analysis of variance, much attention has been paid to non-constancy and non-normality of the "error" contribution. (The recent papers in *Biometrics* by Eisenhart [4], Cochran [3] and Bartlett [1] discuss these matters and give references.) The present writer is usually much more concerned with and worried about non-additivity, and until recently has suffered from the lack of a systematic way to seek it out, and then to measure it. (Conversations with Frederick F. Stephan have contributed greatly to this development and presentation.)

The purpose of the present paper is to indicate such a way, when the data is in the form of a row-by-column table. (The professional practitioner of the analysis of variance will have no difficulty in extending the process to more complex designs.) We shall show how to isolate one degree of freedom from the "residue", "error", "interaction" or "discrepance", call it what you will. There are two known situations to which this single degree of freedom is expected to react by swelling:

- (1) when one or more observations are unusually discrepant;
- (2) when the analysis has been conducted in terms where the effects of rows and columns are not additive.

The first situation is quite familiar and requires little explanation. The second occurs often enough, but may not be noticed. An example may help to fix the ideas.

Let us construct an artificial example with 3 rows and 4 columns, with each entry contributed to overall, by rows, by columns, and by cells. Suppose that these contributions are as follows:

*Prepared in connection with research sponsored by the Office of Naval Research.

†Presented to the Biometrics Section and the Biometric Society at Cleveland, December 29, 1948.

in general	by rows				by columns				by cells			
1 1 1 1	4	4	4	4	6	1	-4	0	1	-2	1	0
1 1 1 1	-3	-3	-3	-3	6	1	-4	0	0	-1	2	-3
1 1 1 1	-3	-3	-3	-3	6	1	-4	0	0	-2	-1	0

Then the tables and corresponding analyses for the sum of all contributions are:

TABLE 1
ILLUSTRATIVE EXAMPLE IN ORIGINAL TERMS

Values and Means						Analysis of Variance			
12 4 2 5 23 5.8 4 -2 -4 -5 -7 -1.8 4 -3 -7 -2 -8 -2.0							DF	SS	MS
20 -1 -9 -2 8 6.7 -0.3 -3.0 -0.7 0.7						Rows	2	140	70
						Columns	3	157	52
						R × C	6	26	4

Now let us square the entries and divide by 10, rounding to integers. The resulting tables and analyses are:

TABLE 2
ILLUSTRATIVE EXAMPLE IN TERMS OF SQUARES

Values and Means						Analysis of Variance			
14 2 1 2 19 4.8 2 0 2 2 6 1.5 2 1 5 0 8 2.0							DF	SS	MS
18 3 8 4 33 6.0 1.0 2.7 1.3 2.8						Rows	2	24.5	12.2
						Columns	3	46.9	15.6
						R × C	6	84.8	14.1

Notice that all semblance of row or column effects have now vanished, although Table 1 showed large and significant effects. The use of the squared scale has concealed the real effects. (It may be argued that squaring numbers which range from plus to minus is unrealistic. The answer is that this is an extreme example, but one that can be slowly and smoothly changed into a very mild one. There probably is a differ-

ence in degree between this example and what happens in practice, but there is no difference in kind.)

PROCEDURE

How then do we isolate the single degree of freedom? The process is simple, and runs as follows:

- (A) To the row-by-column table, already bordered with sums and means, add a new border of deviations of means from the grand mean (decimal places may be reduced, but the sums of deviations, by rows and by columns *must* be forced to vanish).
- (B) Add an extra column (or row) and enter in each cell the sum of products of the deviations by columns and the entries in its row (or column).
- (C) Accumulate the sum of products between the deviations of row (or column) means and the new entries of (B).
- (D) Calculate the sum of squares of deviations by columns and by rows.
- (E) Divide the square of the number from (C) by the product of the numbers from (D). This is the mean square (and also the sum of squares) for the single degree of freedom.

The process is illustrated on the same example below:

TABLE 3
SAMPLE CALCULATION

					Sums	Means	Deviations	Sums of x-products
	14	2	1	2	19	4.75	2.0	38.4
	2	0	2	2	6	1.50	-1.2	3.6
	2	1	5	0	8	2.00	-0.8	4.6
Sums	18	3	8	4	33		0.0	68.8
Means	6.00	1.00	2.67	1.33		2.75	6.08	
Deviations	3.2	-1.8	0.0	-1.4	0.0	15.44	50.9	

$$(B): 14(3.2) + 2(-1.8) + 1(0.0) + 2(-1.4) = 38.4$$

$$2(3.2) + 0(-1.8) + 2(0.0) + 2(-1.4) = 3.6$$

$$2(3.2) + 1(-1.8) + 5(0.0) + 0(-1.4) = 4.6$$

$$(C): 38.4(2.0) + 3.6(-1.2) + 4.6(-0.8) = 68.8$$

$$(D): (3.2)^2 + (-1.8)^2 + (-0.0)^2 + (1.4)^2 = 15.44$$

$$(2.0)^2 + (-1.2)^2 + (-0.8)^2 = 6.08$$

$$\frac{(68.8)^2}{(15.44)(6.08)} = 50.9.$$

Assigning the mean square 50.9 to the degree of freedom for non-additivity, which is subtracted from " $R \times C$ ", the analysis of variance of Table 2 becomes:

Rows	2	24.5	12.2
Columns	3	46.9	15.6
Non-additivity	1	50.9	50.9
Balance	5	33.9	6.8

Thus the obvious thing about the illustrative example was its non-additivity. The corresponding F value of 7.3 on 1 and 5 degrees of freedom is significant at the 5% level.

EXPLANATION

We have explained what we are looking for—non-additivity—and how to look—last section—but we have not explained what we are really doing. This we shall now try to do. Those experienced with single degrees of freedom may have already recognized the computation as a short-cut method of eliminating the single degree of freedom labeled by

$$\begin{array}{ccccccccc} 6.40 & -3.60 & 0.00 & -2.80 & 2.0 & & & & \\ -3.84 & 2.16 & 0.00 & 1.68 & -1.2 & 3.2 & -1.6 & 0.0 & -1.0 \\ -2.56 & 1.44 & 0.00 & 1.12 & -0.8 & & & & \end{array}$$

where $6.40 = (2.0)(3.2)$, $-3.60 = (-1.8)(2.0)$, $2.16 = (-1.8)(-1.2)$ and so on. We have used the products of the deviations of the row means and the deviations of the column means to label this single degree of freedom. Since the sum of each column and of each row is zero, this degree of freedom is orthogonal to rows and to columns. It must be a part of " $R \times C$ ". This is what we did, but why?

Let us take a special case, where there are row contributions, and column contributions, *and nothing else*. We start with perfect additivity. If x_i is the column contribution (where i goes from 1 to c , the number of columns), and if y_j is the row contribution (where j goes from 1 to r , the number of rows), then the ij entry in the table is

$$a_{ij} = x_i + y_j.$$

Now let us start to analyze a slightly nonlinear function of the a_{ij} . Instead of a_{ij} , consider

$$f_{ij}(a_{ij}) = a_{ij} + \lambda(a_{ij} - a)^2$$

where λ is a small constant, and a is, for convenience, the average $\bar{x} + \bar{y}$ of all the a_{ij} . We find that we can write

$$f_{ij}(a_{ij}) = [x_i + \lambda(x_i - \bar{x})^2] + [y_j + \lambda(y_j - \bar{y})^2] + \lambda(x_i - \bar{x})(y_j - \bar{y}).$$

The first two terms depend, respectively, on the column alone and on the row alone, so the last one contains all the *non-additive* effect due to analysis in terms of $f(a)$ instead of in terms of a . Notice that this non-additive effect is a multiple of

$$(x_i - \bar{x})(y_j - \bar{y}).$$

This means that it occurs in a single degree of freedom, which is identified in terms of $x_i - \bar{x}$ and $y_j - \bar{y}$.

We assumed no error of measurement, or the like, and we wrote $a_{ij} = x_i + y_j$, without an additional term. This means that the difference between the i -th column mean and the grand mean is

$$(x_i - \bar{x}) + \lambda\{(x_i - \bar{x})^2 - (x_i - \bar{x})^2\}$$

which is nearly $x_i - \bar{x}$ when λ is small. Thus a satisfactory approximation to the single degree of freedom we want is that indicated by the coefficients

$$(\text{column mean} - \text{grand mean})(\text{row mean} - \text{grand mean})$$

This is exact for the combination of no error and a very slight change from a to $f(a)$, that is for no error and λ small. This fact plus empirical tests seems enough to warrant recommending general use of this single degree of freedom as a test of non-additivity.

WHAT OF SIGNIFICANCE?

Suppose that the test shows statistically significant evidence of non-linearity—what then? The simplest and laziest thing to do would

be to forget the degree of freedom for non-additivity and go on and use the mean square for the balance in considering for example, the significance of the row effects. *This is not recommended*, for the following reasons:

- (1) In general, results expressed in terms in which effects are additive apply in a broader region and are practically more useful. //
- (2) If the "error" or fluctuating contribution is not normally distributed, then it is not known whether or not the use of the balance mean square unduly inflates the apparent significance of other mean squares (for the case of a normally distributed fluctuating contribution there is no distortion of significance.)

For these reasons, the occurrence of a large non-additivity mean square should lead to consideration of a transformation followed by a new analysis of the transformed variable.

This consideration should include two steps: //

- (a) inquiry whether the non-additivity was due to analysis in the wrong form or to one or more unusually discrepant values;
- (b) in case no unusually discrepant values are found or indicated, inquiry into how much of a transformation is needed to restore additivity. .

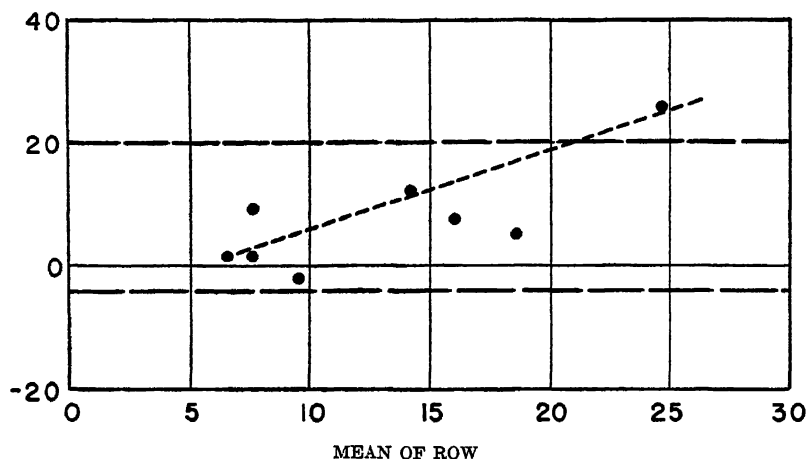
The decision under (a) will depend on an examination of the data and all the background information available in the field—in particular the result of similar inspections of other experiments for non-additivity. What seems to be the best way of inspecting the results of a single experiment so far proposed is to plot the entries in the new column (of sums of cross-products) against the corresponding row means. A single unusually discrepant observation will tend to be reflected by one point high or low and the others distributed around a nearly horizontal regression line. An analysis in the wrong terms will tend to be reflected by a slanting regression line.

The figure shows such a plot, including 2s limits, for

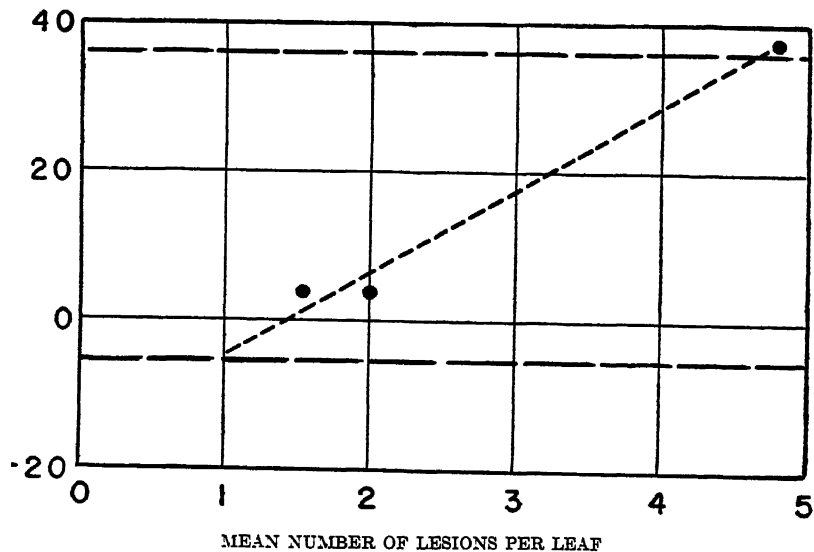
- (A) the illustrative example worked above,
- (B) Youden and Beale's data [6] as simplified by Snedecor [5, p. 44],
- (C) Beall's experiment VI [2] on insect infestation, with plots

GRAPHICAL ANALYSIS OF NONADDITIVITY
(Ordinates are Sums of Cross Products, Dashed Lines are 2S Limits)

A—ILLUSTRATIVE

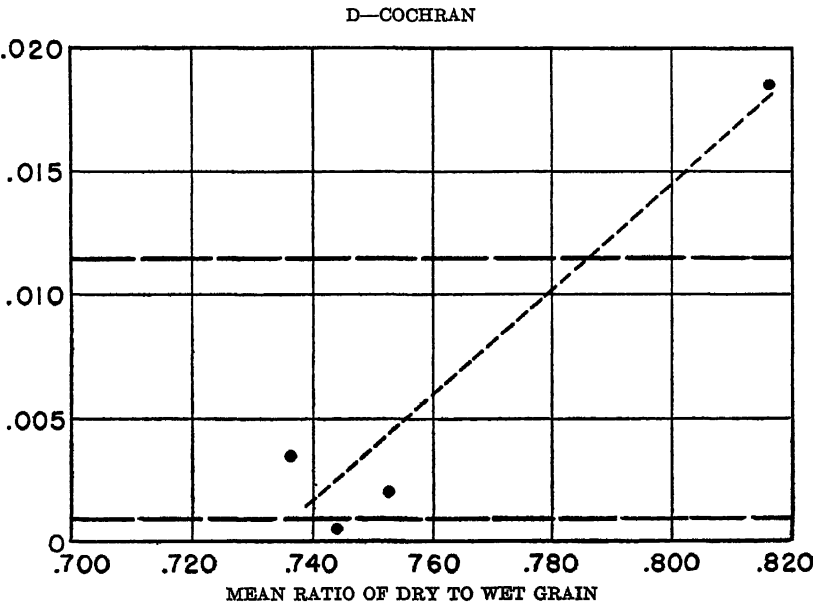
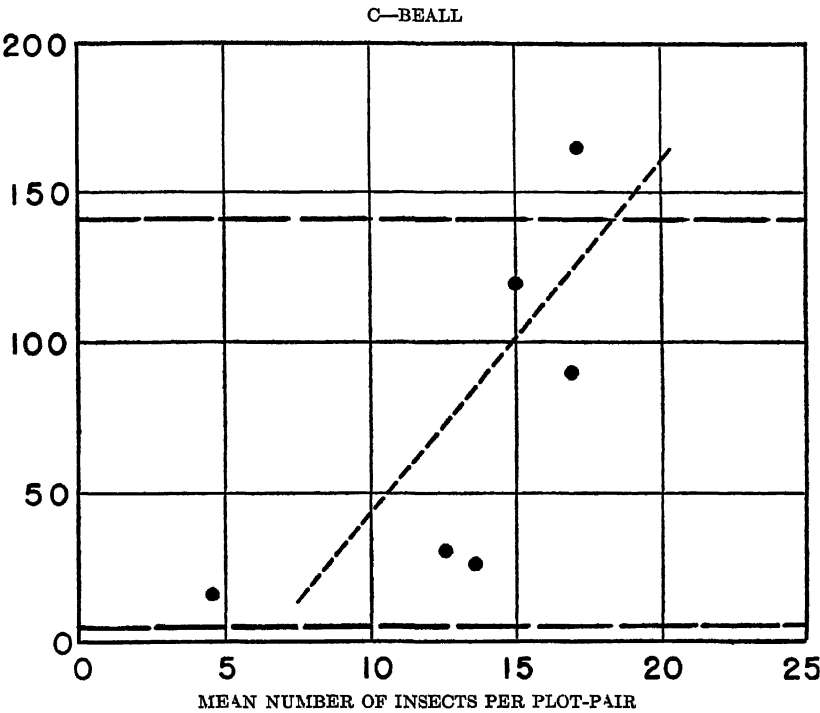


B—YOUDEN & BEALE



treated alike combined (analyzed in terms of numbers of insects).

(D) Cochran's example [3] of an obviously discrepant value.



The limits are set by the formula

$$\left(\begin{array}{c} \text{average} \\ \text{cross product} \end{array} \right) \pm 2 \left(\begin{array}{c} \text{sum of squares of} \\ \text{deviations of column means} \end{array} \right)^{\frac{1}{2}} \left(\begin{array}{c} \text{mean square} \\ \text{for balance} \end{array} \right)^{\frac{1}{2}}$$

For the illustrative example (Case A), this becomes

$$15.53 \pm 2 (15.44)^{\frac{1}{2}} (6.8)^{\frac{1}{2}} = 15.5 \pm 20.5 = -5.0 \text{ and } +36.0.$$

In every one of the four cases, the plotted points could be accounted for by non-additivity due to analysis in incorrect terms. Cases A and D can also be accounted for by a discrepant point. This suggests that it will be hard to make this distinction for single experiments on this scale. When several small experiments are available for analysis, agreement in signs of the slopes of the graphs or equivalently, the signs of the sums obtained in Step C may show up analysis in incorrect terms.

Why does the graph fail to decide about Cases A and D? The reason is simple—either explanation is plausible. If in Case A we alter the upper left-hand entry from 14 to 2, the analysis of variance becomes:

	DF	SS	MS
Rows	2	0.5	0.2
Columns	3	4.9	1.6
Non-additivity	1	0.2	0.2
Balance	5	12.6	2.5

Thus we see that our illustrative table of 3×4 entries *could* have perfectly well come from an additive situation where exactly one entry has been seriously disturbed.

Similarly in Case D, taken from Cochran's paper, if a nonlinear function is chosen so that

$$g(y) = \begin{cases} y, & .704 \leq y \leq .792, \\ .800, & y = 1.035, \end{cases}$$

then his table is converted into one where the F -ratio for non-additivity against balance is 0.8 instead of 27.6. We know that this table arose from an error in computation, but it *could* equally well have come from an additive table analyzed in the wrong terms.

In each case, the graphical solution has gone as far as it reasonably

could in assigning responsibility for the non-additivity. While the graphical analysis is not certain to settle Step (a), it may be expected to be a big help.

AID IN CHOOSING A TRANSFORMATION

If it has been decided that the wrong terms had been used, then the actual size of the mean square for non-additivity must be useful for choosing an appropriate transformation. We lack experience with the more delicate use of such information, so that it seems appropriate to stop here with the following table which shows the connection between the *sign* of the final sum of products (which was +68.8 in the illustrative example) and the type of transformation which may then be appropriate.

TABLE 4
SIGN OF FINAL SUM OF PRODUCTS WHEN CERTAIN TRANSFORMATIONS
ARE APPROPRIATE (VALUES OF x OR $x + a$ NON-NEGATIVE)

Transformed values which are additive*	Conditions needed	Sign when x is analyzed	Important special cases
x^p or $(x + a)^p$	$0 \leq p < 1$	+	$\sqrt{x}, \sqrt{x+1}$
	$p = 1$	0	(x)
	$1 < p$	-	x^2, x^3
$\log (x + a)$	(none)	-	$\log x, \log (1 + x)$

*Multiplication by a fixed constant and addition or subtraction of a fixed constant freely possible

While the removal of non-additivity by transformation usually tends to stabilize the variance, there may be cases where the variance is notably non-constant after transformation. In such cases, analysis of the transformed data using weights seems appropriate.

APPENDIX

VALIDITY OF THE ANALYSIS

This section is prepared for those who may feel that the method of obtaining the "single degree of freedom" may not produce quantities with the usual distribution.

The basic fact is this: If $u_1, u_2, \dots, u_k; v_1, v_2, \dots, v_m$ have some

joint distribution, and if, for fixed u_1, u_2, \dots, u_k , the distribution of v_1, v_2, \dots, v_m exists and is *always the same*, then the marginal distribution of v_1, v_2, \dots, v_m exists and, indeed, is the same, and, furthermore, u_1, u_2, \dots, u_k and v_1, v_2, \dots, v_m are independent. This can be established either by general considerations or by analytical detail.

To apply this in our case, let u_1, u_2, \dots, u_k be the row and column means, and let v_1 and v_2 be the sums of squares for non-additivity and for the balance. If the situation is additive, and the cell effects are normally distributed, and u_1, u_2, \dots, u_k are fixed, then v_1 and v_2 are independently distributed like σ^2 times chi-squares on 1 and $rc - r - c$ degrees of freedom. Hence v_1 and v_2 have these distributions, and are independent of all functions of row and column means. Thus the F -tests of rows, columns, or non-additivity against balance are valid.

In the presence of non-additivity and/or non-normality, the usual arguments indicate that the F -test is, if anything, conservative.

REFERENCES

- [1] Bartlett, Maurice S. The Use of Transformations. *Biometrics* 3, 39-57, 1947.
- [2] Beall, Geoffrey. The Transformation of Data from Entomological Field Experiments so that the Analysis of Variance becomes Applicable. *Biometrika* 32, 243-262, 1942.
- [3] Cochran, W. G. Some Consequences when the Assumptions for the Analysis of Variance are not Satisfied. *Biometrics* 3, 22-38, 1947.
- [4] Eisenhart, Churchill. The Assumptions underlying the Analysis of Variance. *Biometrics* 3, 1-21, 1947.
- [5] Snedecor, George. *Statistical Methods*. The Collegiate Press, Ames, Iowa; 4th edition, 1947.
- [6] Youden, W. J. and Beale, Helen Purdy. A Statistical Study of the Local Lesion Method for Estimating Tobacco Mosaic Virus. *Contributions from the Bouce Thompson Institute* 6, 437-454, 1934.

ON A STATISTICAL APPROXIMATION TO THE INFECTION INTERVAL

J. B. CHASSAN*

IN A PREVIOUS PAPER (2) the existence of strong correlation between the logarithms of the morbidity rates of a group of respiratory diseases for successive calendar month-pairs was demonstrated. The case rates involved pertain to the combined incidence of catarrhal bronchitis, acute coryza, acute catarrhal pharyngitis and laryngitis, and influenza, as diagnosed and reported in the United States Army. Where C_i is the case rate observed in the i -th calendar month, and C_{i+1} , the corresponding rate observed in the succeeding calendar month of the same year (or the same winter when $i = \text{December}$), the value of $r_{\log C_i, \log C_{i+1}}$ for the twelve month-pairs averaged .84, each of the twelve coefficients being based upon some 38 observations, according to the number of years for which data were available for each month-pair. The purpose of the present paper is to relate some of the results obtained in connection with ref. (2) to the *law of mass action in epidemiology*, and to derive therefrom an estimate of the infection interval for an assumed period of immunity following infection, or conversely, an estimate of the period of immunity corresponding to a known infection interval. In connection with the actual numerical values presented, it should be noted that they pertain to a *group* of diseases and therefore can be interpreted only as average for the group as a whole.

The law of mass action in epidemiology states that the rate at which a contagious or epidemic disease spreads in a community is proportional to the product of the number of infectious individuals and the number of susceptibles in the community. If two consecutive time intervals are chosen such that the length of each interval is equal to the period between contact and case manifestation (i.e., the incubation period), a contact between an infectious person and a susceptible in the first interval will result in a new case during the second. Then the law of mass action may be written as

*The author wishes to acknowledge the helpful criticism and suggestions of Prof. John W. Tukey of Princeton University in connection with this paper.

$$C_{i+1} = \frac{1}{m} S_i I_i \quad (1)$$

in which

- (a) C_{i+1} is the expected number of cases (or the case rate) during the $(i + 1)$ -th period.
- (b) S_i is the average number of susceptibles in the i -th period.
- (c) I_i represents the average number of infectious individuals during the i -th period.
- (d) m^{-1} is the proportionality constant reflecting such factors as the degree of crowding in a community, seasonality; more abstractly "infective power".

For the case in which the period of communicability following infection is relatively short, it is convenient to consider incidence in successive intervals whose lengths are each equivalent to the infection interval, rather than to the incubation period. The infection interval may be defined as the average period between the manifestations of two cases, one case resulting from contact with the other. It can be regarded as the sum of two components: first, the (average) time it takes for adequate contact to take place between a newly infected person and a susceptible, and then, the period between contact and manifestation. In such a case, we may replace I_i by C_i in equation (1), obtaining Soper's formula,

$$C_{i+1} = \frac{1}{m} S_i C_i \quad (2)$$

which gives the relationship between incidence rates in two consecutive periods whose lengths are each equal to that of the infection interval.

Soper (1) has also stated the relationship for the case in which the incidence rates are taken over successive periods of arbitrary length. If C_i is the case rate observed during the i -th month, and S_i is the average number of susceptibles in the i -th month, then

$$C_{i+1} = \left(\frac{S_i}{m} \right)^p C_i \quad (3)$$

where C_{i+1} is the incidence rate in the $(i + 1)$ -th month, and p represents the number of infection intervals in one month.

If C_i is expressed as a daily incidence rate in terms of the number infected per day out of each 1000 population, and if n is the number of days of immunity following infection, then nC_i will give the average number per 1000 population who are not susceptible, by virtue of recent infection, during the month in which C_i is observed. On the assumption

of general susceptibility in the population, the corresponding number of susceptibles per 1000 will then be given by

$$S_i = 1000 - nC_i \quad (4)$$

Substituting this value in (3), we obtain

$$C_{i+1} = \left(\frac{1}{m}\right)^p (1000 - nC_i)^p C_i \quad \text{or}$$

$$\log C_{i+1} = p \log m^{-1} + p \log (1000 - nC_i) + \log C_i \quad (5)$$

Interpreting this equation in a statistical sense, i.e., as a regression function in which $x_{i+1} = \log C_{i+1}$ is regarded as the average value corresponding to a fixed observation of $x_i = \log C_i$, the data for the group of respiratory diseases under consideration indicates that the true regression curve of x_{i+1} on x_i increases monotonically with slowly declining slope over the actual range of observations. Apart from sampling differences a straight line of the form

$$x_{i+1} = a + bx_i \quad (6)$$

fitted by the method of least squares should lie close to the regression curve over the range of observed values of x_i , and the slope of the line, b , should very nearly equal the slope, β , of the secant which intersects the true regression curve at points corresponding to the lowest and highest of the observed values of x_i , respectively. An approximation to the infection interval can then be obtained by equating b , the slope of the linear regression of x_{i+1} on x_i , to β , the slope of the secant.

If C_{i_L} represents the lowest of the observed rates in the i -th month, and C_{i_u} , the highest; their substitution, in turn, for C_i in equation (5) above yield, as co-ordinates of the secant at the points of intersection with the mass action curve, the points,

$$\left(\log C_{i_L}, \log \left\{ \left(\frac{1}{m} \right)^p (1000 - nC_{i_L})^p C_{i_L} \right\} \right)$$

and

$$\left(\log C_{i_u}, \log \left\{ \left(\frac{1}{m} \right)^p (1000 - nC_{i_u})^p C_{i_u} \right\} \right)$$

respectively, where each ordinate is expressed as a function of the corresponding abscissa.

Then, from elementary analytic geometry the slope of the secant will be

$$\beta = 1 + p \left(\frac{\log \left(\frac{1000 - nC_{i,u}}{1000 - nC_{i,l}} \right)}{\log (C_{i,u} / C_{i,l})} \right)$$

Solving for p , and substituting b for β ,

$$p = (b - 1) \left\{ \frac{\log (C_{i,u} / C_{i,l})}{\log \left(\frac{1000 - nC_{i,u}}{1000 - nC_{i,l}} \right)} \right\} \quad (8)$$

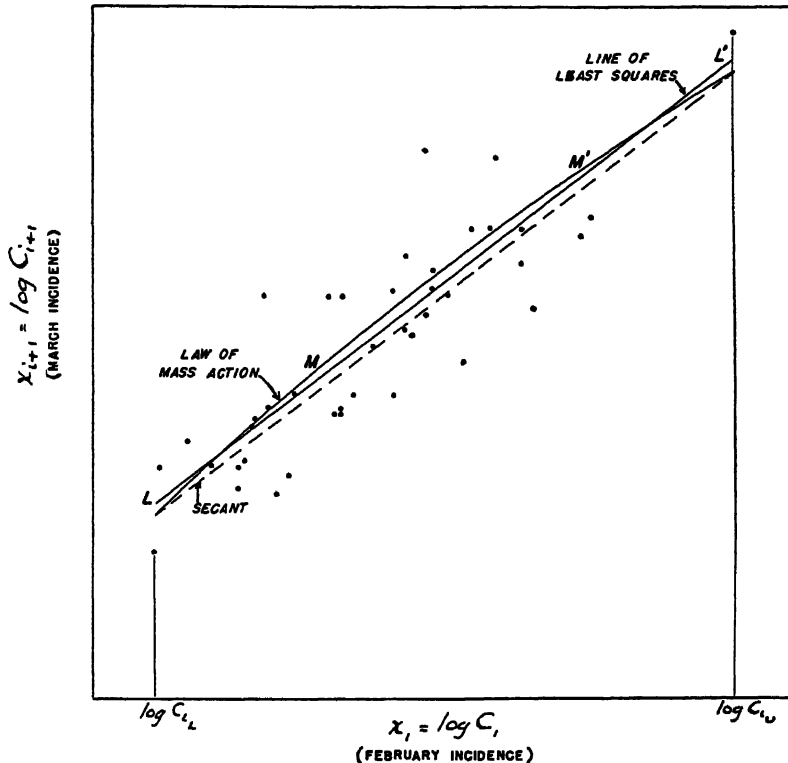
Upon applying formula (8) to the data of reference (2) for the twelve month-pairs, a median value of $p = 15$ was obtained on the assumption of three weeks of incidence as equivalent to the number of non-susceptibles, i.e., when $n = 21$. Since the incidence data were taken over monthly intervals, the corresponding estimate of the average infection interval is 2.0 days. On the assumption that $n = 28$, the median value of p is 11, and the infection interval, 2.8 days. Finally, if the assumption is made that $n = 42$, an average infection interval of 4.1 days is estimated. Thus in the neighborhood of the assumed values of n , the ratio of the period of immunity to the length of the infection interval is approximately 10 : 1.

Illustrating the procedure graphically, the chart given shows:

- (a) a plotting of observed points corresponding to the February-March relationship
- (b) a theoretical drawing of (5), represented by the curve MM' , and interpreted as a regression curve
- (c) the least squares linear regression of x_{i+1} on x_i , LL' , fitted to the scatter of points
- (d) a secant to the curve MM' , drawn as a dashed line; the secant is drawn so that it intersects the curve to the left at the point whose abscissa is $\log C_{i,l}$, where $C_{i,l}$ is the smallest of the observed values C_i , and to the right, at the point whose abscissa is $\log C_{i,u}$, where $C_{i,u}$ is the largest of the observed values of C_i .

The position of the curve MM' in relation to its secant and to the least squares line LL' , (again, apart from sampling errors) can be determined by formulating the vertical distance between MM' and the secant. By differentiation, both the maximum distance and the value of $\log C_i$ at which the maximum distance occurs can be determined. Thus if the equation of the secant is given by

$$\log C_{i+1} = \alpha + \beta \log C_i, \quad (9)$$

GRAPHICAL REPRESENTATION OF THE ESTIMATING RELATIONSHIPS IN
 THE APPROXIMATION TO THE INFECTION INTERVAL


THE SECANT TO THE LAW OF MASS ACTION CURVE, INTERSECTING THE CURVE AT EXTREMES OF x_L , IS ASSUMED PARALLEL TO THE LEAST SQUARES LINE OF REGRESSION.

it will be found, by substituting $C_{iL} = C_i$ in equation (9) and in (5) that

$$\alpha = p \log m^{-1} + p \log (1000 - nC_{iL}) + (1 - \beta) \log C_{iL}.$$

The distance from the secant to the curve will then be

$$\phi = p \log \left(\frac{1000 - nC_i}{1000 - nC_{iL}} \right) + (1 - \beta) \log \frac{C_i}{C_{iL}} \quad (10)$$

where b can be substituted for β , and p is obtained from (8).

The maximum value of ϕ can, of course, be obtained by differentiation with respect to C_i , or $\log C_i$, and equating to zero. Then the curve

MM' can be closely approximated from the position of the least squares line. Taking

$$M = L + \phi - 1/2 \max \phi$$

where, for a fixed value of $\log C_.$, L is the corresponding value of $\log C_{.+1}$ on the least squares line, and ϕ is taken from (10), M is the corresponding value of $\log C_{.+1}$, on the curve.

For an assumed value of p , equation (8) can be solved for n . From

$$\left(\frac{C_{.u}}{C_{.L}}\right)^{(b-1)/p} = \frac{1000 - nC_{.u}}{1000 - nC_{.L}}$$

we obtain

$$n = \left\{ \frac{1000 \left[1 - \left(\frac{C_{.u}}{C_{.L}} \right)^{(b-1)/p} \right]}{C_{.u} - C_{.L} \left(\frac{C_{.u}}{C_{.L}} \right)^{(b-1)/p}} \right\}$$

In applying the foregoing type of analysis the following modifications or limitations should be considered:

(i). We have assumed that for a fixed month-pair the infectivity factor, m^{-1} , is constant, except for random variation. Were it not for the fact of a declining number of susceptibles, $S_.$, with increasing $C_.$, as described by equation (4) above (i.e. if $S_.$ were constant over the range of $C_.$), the mass action curve as given by equation (5) above would assume linear form with slope unity. But the declining value of $S_.$ has the effect of causing the slope to drop with increasing $C_.$, so that apart from sampling errors, the slope of b (and of β) will be less than unity. This can be seen quite easily if we write equation (3) as

$$C_{.+1} = A.C_.$$

Then, if $A_.$ were constant for all $C_.$, a plotting of the curve (on a log-log scale) would yield a straight line parallel to

$$C_{.+1} = C_.$$

at a vertical distance of $\log A_.$ But with the damping effect of the decline of susceptibles as $C_.$ increases, $A_.$ correspondingly decreases; and if, for example, $\log A_.$ is still positive the distance between the two lines decreases with increasing $C_.$, and it then follows that $\beta < 1$. The same result would, of course, apply when $\log A_.$ is negative.

Now if the situation were such that as $C_.$ increases various preventive measures are taken which significantly reduce the infectivity factor,

further damping will take place, and b will become smaller. To take this into account it would then be necessary to adjust upward the value of b , resulting in a corresponding increase in the length of the infection interval for each of the assumed values of n .

(ii). Equation (4) above implies that the entire population is potentially susceptible, and that the only immunes present at any given time, are those individuals who have gained immunity for a short period by virtue of recent infection. If, however, only a fraction, q , of the entire population are potentially susceptible, then instead of (4), it would be necessary to write

$$S_t = 1000q - nC_t \quad (11)$$

and substituting this, instead of (4) in (5), and in (7) and (8), it will be seen that for the same observed value of b , a somewhat longer infection interval would be estimated, depending on the degree of departure of q from unity.

(iii). Equations (4) and (11) will progressively lose accuracy as n gets very large. Thus, if the period of immunity were to last several months, these expressions would require modification to take account of variation in C_{t-1} , C_{t-2}

References (3) and (4) listed below, and others listed in these references, discuss various aspects of the law of mass action of importance in connection with epidemic theory.

REFERENCES

1. Soper, H. E. The Interpretation of Periodicity in Disease Prevalence. *Jour. Roy. Statist. Soc.* 92, 34-61, 1929.
2. Chassin, J. B. The Autocorrelation Approach to the Analysis of the Incidence of Communicable Diseases. *Human Biology* 20, 2, 90-108, 1948.
3. Wilson, E. B. and Worcester, Jane. The Law of Mass Action in Epidemiology. *Proc. Nat'l Acad. Science* 31, 1, 24-34, 1945.
4. Wilson, E. B. and Burke, Mary H. The Epidemic Curve. *Proc. Nat'l Acad. Science* 28, 9, 361-367, 1942.

QUERIES

70 **QUERY:** I am carrying forward research on little known or on unknown tropical feedstuffs. For this research, rats, baby chicks and pigs are being employed. The unknown feedstuffs are evaluated singly and in combinations. I would appreciate your opinion on the proper method of statistical analysis for our data.

As an example and for brevity, here are some actual data from a pilot trial, together with the analysis of variance.

WEIGHT GAINS OF BABY CHICKS

No. chicks	Treatment				Entire sample
	1	2	3	4	
1	55	61	42	169	
2	49	112	97	137	
3	42	30	81	169	
4	21	89	95	85	
5	52	63	92	154	
	219	355	407	714	1695

ANALYSIS OF VARIANCE

Sources	D.F.	S.S.	M.S.
Lot means	3	26235	8745**
Individual	16	11559	722
Total	19	37794	

The *F*-test in the above case is highly significant indicating that we are not dealing with a single population. This method of analysis however does not provide us with a means of stating that treatment No. 3 is better than No. 1 or No. 2 is better than No. 4, etc. Could you provide us with the most valid method with which we could make these comparisons?

ANSWER: Happily this perennial question has been provided with an answer by Dr. John W. Tukey in the June issue of this Journal (Vol. 5: pages 99-114, 1949). Tukey's method indicates a gap between the first three treatments and the fourth. At a risk of less than one per hundred, one would reject the hypothesis of no difference between treatments No. 3 and No. 4.

There is not sufficient evidence to cut off the straggling mean of treatment 1 ($P = 0.17$). Finally, applying the F -test as indicated by Tukey, one does not reject the hypothesis that lots 1, 2, 3 are drawn from a common population ($P = 0.1$).

I assume that your experiment was conducted so that environmental differences were randomly distributed over all the chicks in the experiment; otherwise, there is no unambiguous answer to the question about the effects of treatments.

QUERY: In an experiment in which one half of the controls
71 reacted positively and one half negatively, it would seem that chi-square should be the same whether one uses the formula,

$$\chi^2 = 2(x - m)^2/m,$$

or the formula for the 2×2 table,

$$\chi^2 = \frac{(ad - bc)^2(a + b + c + d)}{(a + b)(c + d)(a + c)(b + d)}$$

But this is not the case. Why?

For example, suppose 200 animals are divided equally among experimentals and controls. Then, according to the proposition under consideration, suppose 50 controls live and 50 die, and suppose 63 of the experimentals live and 37 die. Is the experimental procedure effective?

By the 2×2 table, $\chi^2 = 3.438$, not significant. But by the other formula, comparing the experimentals with a 1 : 1 ratio, $\chi^2 = 6.760$, highly significant. Why do not the two methods agree?

ANSWER: You have described two different experiments leading quite properly to different values of chi-square. In the first experiment there are only 100 animals, all treated experimentally. The assumption is made that in the untreated population the ratio of the numbers living and dying is 1 : 1. The hypothesis being tested is that the same ratio applies to the treated population; that is, that the treatment is without effect. The value of chi-square, 6.760,

would lead to rejection of the hypothesis with P approximately 0.01. In this experiment there are no controls because the experimenter supplies the information about how controls behave.

The second experiment contains 200 animals, but half of them are used to get evidence about the behavior of the untreated population. Here the experimenter either has no knowledge of the behavior of the controls or is unwilling to rely on his knowledge. In this experiment, the hypothesis being tested is that the experimentals and controls have the same ratio, but the value of the ratio is not specified. The experimenter supplies less information than he did in the first experiment. The result is that the same number of experimentals, divided in the same ratio, lead to less certainty about the conclusion.

Querist feels that the chance division of the controls in the 1 : 1 ratio is equivalent to the 1 : 1 hypothesis which was set up in the first experiment. That this is not true may be clear if he considers the 95 percent confidence interval based on a sample of 100 equally divided in outcome. This interval is from 40 percent to 60 percent. The corresponding 99 percent interval is from 37 percent to 63 percent. Evidently the information supplied by such a sample of controls is far less than that furnished by the experimenter in postulating the 1 : 1 ratio for the population of controls.

72 QUERY: Hace un tiempo, se discutía en una reunión efectuada entre técnicos especialistas en maíz las exigencias para aprobar un híbrido o rechazarlo.—

Alguien sugirió aceptarlos cuando los rendimientos eran estadísticamente significativos.—

Y aquí comenzó la controversia. Otro técnico tomó la palabra para exponer su pensamiento al respecto. Dijo, que si se efectuaba un ensayo con todo cuidado, las exigencias para considerar un determinado híbrido estadísticamente superior a otro (altamente significativo), serían muy reducidos. Por ejemplo, un 3% de diferencias en los resultados, podría ser lo suficiente para que de acuerdo al análisis estadístico, se considere a un híbrido superior.—

Esto llevaría a un error, pues un 3%, en la práctica (en el gran cultivo) no tendría ninguna importancia, por lo que el procedimiento era erróneo. En cambio se mostró partidario de exigir un 10% de diferencia en los rendimientos y fijar un error standard de por ejemplo 6%.—

Desde luego, no sé a ciencia cierta quien tiene razón, por lo que recurro a Ud. a fin de que me evacúe la consulta. Puede hacerlo en inglés.—

Yield trials of various crops are usually conducted for one
ANSWER: of two reasons: (1) to provide a test for a particular hypothesis or (2) to provide information which can be used as a guide in making recommendations over a range of soil and climatic conditions.

In the first instance an efficient experimental design and adequate replication are necessary so that the desired tests may be performed with the required precision. The number of replications and choice of design will, in part, be dictated by past experience as to soil variability, etc.

In the more general case where yield trials are conducted to provide information which will serve as a guide in making general recommendations, the situation is quite different. It is well established that different varieties respond differently in different years and at different locations. Therefore, varietal trials must be grown at several locations and in different years. Thus, there is little point in striving for "statistical significance" in each of the individual tests. An increase in number of replications for any single test will have little effect in reducing the magnitude of the variety \times year or variety \times location interaction.

The general practice in yield trials is not to select one or a few of the apparently superior items, but rather to discard a group of the poorer items. The items remaining are then tested further to provide additional information on performance.

If a number of varieties are tested over a series of years and locations, the outcome will almost certainly be a group of varieties which are so similar in yield and other characteristics that the differences among them will not be statistically significant. The best estimates of the relative value of the varieties in this group will be the actual averages obtained.

G. F. SPRAGUE

THE BIOMETRIC SOCIETY

By the time this number of *Biometrics* reaches you, each member of the Society will have received his free copy of our first Directory. Additional copies have been printed to send to new members as they are enrolled. It is available to non-members for 50 cents. Until a new edition is warranted, we propose issuing an annual supplement. As you will have discovered, the Directory includes a list of officers, the constitution of the Society, the Council by-laws, and the statutes of each region as well as the alphabetical membership list and a geographical summary. The information provided for each member includes his professional connection as recorded in the Secretary's office on June 15 and his major field of interest. Later, we hope to summarize the distribution of members among the different fields of interest. Although the Society has been in existence for less than two years, the geographical breakdown shows that we had 888 members in 33 different countries when the Directory went to press. The first and largest organized region was the Eastern North American, with 478 members. The other regions in order of formation were the British with 111 members, Western North American with 73 members, Australasian with 37 members, Indian with 43 members and French with 47 members. In addition, there were 99 members-at-large.

Since the last issue, the Council has approved the statutes of the Australasian, Indian and French Regions. These are already included in the Directory, so that they need not be reprinted here.

Developments in France are of unusual interest. The biometricians there have adopted a dual organizational plan in accord with a law of 1901 governing official French societies. They have formed the autonomous Société Française de Biométrie. At the same time they have formed the Région Française of the Biometric Society and provided that all full members of the Société Française de Biométrie shall be members of the Biometric Society. In view of this interesting development the tentative proposal of a joint French-Italian region has been abandoned. At the last meeting of the Société Française, on May 17 at the Laboratoire de Zoologie de la Faculté des Sciences, Paris, the following communications were presented: "La rehabilitation de l'homme moyen" by

M. Frechet, "Facteurs lateraux et facteurs sexuels dans la morphologie des empreintes digitales" by R. Turpin and M. P. Schutzenberger, and "Etudes biometriques sur le colibacille" by J. Dufrenoy.

Within the last months the following regional officers have been elected and confirmed by Council: British Region: Vice-President, J. W. Trevan; Secretary, D. J. Finney; Treasurer, K. Mather; Regional Committee, J. O. Irwin, J. I. M. Jones. Indian Region: Vice-President, P. C. Mahalanobis; Secretary, C. Radhakrishna Rao; Treasurer, Mohanlal Ganguli; Regional Committee, V. M. Dandekar, K. Kishen, K. R. Nair, U. S. Nair, V. G. Panse, P. B. Patnaik, B. Ramamurthy, R. V. Sukhatme, V. D. Thawani.

Since last November the Society has been provided with temporary headquarters in a pleasant room at 321 Congress Avenue in New Haven by the Department of Public Health of the Yale University Medical School. This room, however, will be required for new activities in the next academic year. Through the kindness of the Department of Applied Physiology, the Society has had the good fortune of obtaining a larger room at 52 Hillhouse Avenue in the main part of the University, and moved there on July 5. We would be very glad to welcome any visiting members at our new headquarters. We are very sorry to lose the services of Mrs. Elizabeth Weinman, who was Executive Assistant to the Secretary through June 30. The Society has benefited greatly from her efficient handling of the many details of the Secretary's office and wishes her well in her new undertaking. We have been fortunate in obtaining as her successor Mrs. Irving N. Fisher, who knows at first hand all of the countries where we have regions and most of the other countries where we have members.

NEWS AND NOTES

At the Raleigh branch of the Institute of Statistics there is a small news publication called the "Leaky Gasjet" which is printed irregularly depending upon the quantity of choice gossip acquired by its faithful seekers. The following excerpt was taken from the June, 1949, edition.

Dear Gasjet Editor:

I am a newly created Ph.D. in Experimental Statistics and I am worried because I expect to do consultation and I am afraid that the research workers will ask me questions that I won't be able to answer. What shall I do?

Phidler

Dear Phidler:

Here are a few simple devices which should prove useful to you in your consulting work. Relax, once you have mastered them you have absolutely nothing to worry about.

Research Worker: *Confidently.* I have done an experiment, Mr. Phidler, in which I have two plants, one of each variety, in each pot and fifteen pots. Can you tell me how to analyze it so as to show that Variety A is taller than Variety B? I realize *laughing selfconsciously* that this is a very elementary question but . . .

Phidler: *Frowning.* *Naturally as a new Ph.D. this is far too difficult a question for him, but he is not alarmed.* Just what do you mean by taller?

This illustrates both the Device of the Counterquestion and the Device of the Definition of Terms.

Research Worker: *A bit taken aback.* Taller? Well I mean bigger—not not bigger—

Phidler: *Sternly.* Come now, we cannot get anywhere unless we have specific, operational definitions.

Research Worker: Yes, of course. What I meant was I measured the height of each plant and—

Phidler: The external or the internal height? *He pauses, but Research Worker is unable to answer.* A similar problem came up in the Jour.-Roy-Stat-Soc-Supple-eleventy two-page 476.

The Device of the Non-Existent Reference

Research Worker: *Awed.* What was that reference again?

Phidler: No matter. It's by Gregory Hairshirt. I knew Hairshirt in kindergarten—an idiot—his papers were demolished by Smirkley Annals of Applied Human Genetics. Let's get back to our little problem.

The Device of Complete Familiarity with Everyone and Everything

Research Worker: *Relieved.* Yes, Yes. Now I thought this design—

Phidler: Design? *Laughs* Yes **design**. You realize of course that you should have used a cuboidal lattice in this experiment.

The Device of the Wrong Design

Research Worker: I—well I didn't know—

Phidler: *Aloud to the walls.* How do these research workers expect us to get anything out of their data when they use any old design. Ah well, I suppose we can work it out by matrix methods. Tell me, what is the Cost Function for height in this problem?

The Device of the Unnecessary Complication

Research Worker: Cost? I don't know—I thought this was a simple *sob!* problem—but after all I'm only a miserable research worker and not a statistician *alas!*

Phidler. *Bcnevolently.* Now, now, don't cry. I will help you. This is really a very simple problem.

The Device of Reversing your Field

Research Worker. *On his knees.* For you, perhaps, O Master. *The research worker is now in the proper frame of mind for consultation. From here on in Phidler can do ANYTHING.*

AFRICA—Among our new members, **Henri Marchand**, Dakar, Sénégal, West Africa, writes, "My researches are purely theoretical in the field of mathematical genetics. As soon as my present studies on the part that a single body can have on the evolution of a population are advanced, it will give me great pleasure to send you a report on the results at which I will have arrived."

AUSTRALIA—**Helen Turner** had plans all completed to attend the Second International Biometrics Conference in Geneva and to spend six months in Cambridge. Unfortunately family illness has intervened and the trip has been postponed. **D. B. Duncan** is busy developing a teaching program in Statistical Methods in the University of Sydney. Three new courses have been set up and the first graduate in Agricultural Science with Honors in Statistical Methods, **J. A. Morris**, took his degree this March and is now working in animal genetics in the Division of Animal Health and Production of C.S.I.R.O. **H. O. Lancaster** of the Commonwealth Health Department has just completed a year's study in England and is now on his way back to Australia. **E. A. Cornish** has a new F_1 to carry on the statistical tradition. **C. W. Emmens**, author of the recently published *Principle of Biological Assay*, is coping well with a large demand for presentation of papers to scientific societies in Sydney.

FINLAND—**Leo Törnqvist**, Chairman of the Institute of Statistics, University of Helsinki sends a brief note. He writes, "The Institute of Statistics in the University of Helsinki was founded in 1945, but has started its activity only in 1947. Its Chairman is the professor in statistics of the University, and an M.A. works there as assistant. The Institute is partly a statistical library, partly an advisory and direction office for the students of statistics. In addition the chairman, the assistant, and the more progressed students work with special statistical researches for outsiders. The received tasks have chiefly been from the branches of population—prognostics and analysis of economic time-series. The teaching in statistics belongs in the University under the Faculty of the Political Sciences. The student can choose statistics in the M.A.-examination for his chief subject or for one of his side subjects. After the M.A.-examination it is possible to go on with the studies as far as to the doctoral thesis. My special interests in statistics are the theoretical and economical problems."

INDIA—**D. N. Nanda** has taken up the position of a Statistician for Indian Army Ordnance Corps. He writes, "In this capacity I am to conduct Applied Research on the following subjects: (1) Design and Analysis of Experiments, (2) Quality Control, (3) Sampling Surveys (including inspection methods). There are a number of other topics on which I may have to work from time to time." He would appreciate being informed of the latest developments in these fields.

UNITED STATES—On February 1, **Alexander G. Ruthven**, president of the University of Michigan, announced the establishment of the Institute for Social Research. "The institute will be directed by **Rensis Likert** and will provide a unified administration for two units already existing at the University, the Survey Research Center and the Research Center for Group Dynamics. **Angus Campbell** will succeed Mr. Likert as Director of the Survey Research Center, which will continue its major programs of research in such fields as: studies of economic behavior and motivation; studies in human relations and organization; studies of the American public's understanding of major national and international issues; and the development of sampling survey methodology. **Dorwin Cartwright** will continue as Director of the Research Center for Group Dynamics. As a part of the Institute for Social Research this group will continue its program of research on the factors influencing productive and harmonious group functioning. It will continue its studies on human relations in industry, leadership, communication within groups, inter-group relations, and the social satisfaction of community life. As a result of the joining of the two centers, the Institute is better able to bring to bear quantitative and experimental research methods on complex and important social problems. Research findings of the Institute are communicated not only through teaching and scientific publications, but also through consultation and training in various organizations. The staff of the Institute includes over 350 persons engaged in full time or part time work. Approximately 125 of this number are located in Ann Arbor. Although most of the professional staff are social psychologists, various other social sciences are represented." **Melville A. Taff, Jr.**, formerly with the Louisiana State Department of Health, New Orleans, is now with the Territory of Hawaii Department of Health, Honolulu, as Chief of the Bureau of Health Statistics. Mr. Taff writes, "The Bureau is being expanded to provide statistical service for the entire department. Additional tabulating equipment has been ordered and more statistical personnel will be added as necessary. A central statistical service unit is the prime objective. An Act patterned after the Uniform Vital Statistics Act was passed at the 1949 session of the Legislature and now awaits the signature of the Governor. Once signed one of the first moves will be to consolidate small and sparsely populated registration districts and wherever possible and practical to appoint the local health officer as local registrar. Office methodologies are being reviewed and revised procedures are being written." **Paul T. Bruyere** formerly with the Army Institute of Pathology is now with the Division of Tuberculosis, United States Public Health Service. He, with **Martha**

Bruyere, is making a study of the early development of tuberculosis among student nurses. **Jack Chassan** also joined the United States Public Health Service and is working with the Bruyere's on the student nurse study. He recently left the Office of the Surgeon General, Department of the Army. **Allen B. Burdick** is now Assistant Professor of Agronomy, University of Arkansas, Fayetteville. He is initiating research in the development of grain and forage types of sorghum and will teach a course in the genetics of plant breeding. His theoretical research will continue to emphasize the mathematical aspects of quantitative inheritance. Mr. Burdick was with the Atomic Energy Commission at the Genetics Division, University of California, Berkeley. **H. M. C. Luykx** is resigning his position as Associate Professor of Preventive Medicine, at New York University College of Medicine, to accept appointment as Biometrician for the Atomic Bomb Casualty Commission in Japan. The Commission operates under the Committee on Atomic Casualties of the National Research Council, Washington, by directive of the President, and is sponsored by the Atomic Energy Commission. Mr. Luykx will be stationed in Japan for about two years, where he will make his home in Kure, with frequent visits to Hiroshima and Nagasaki. **R. L. Murphree** recently resigned his position with the Bureau of Dairy Industry at Jeanerette, Louisiana, to accept a position as Associate Professor of Animal Husbandry at the University of Tennessee. **Kenneth S. Cole**, formerly with the Institute of Radiology and Biophysics of the University of Chicago, is now Scientific Director of the Naval Medical Research Institute at Bethesda, Maryland. **Theodore A. Bancroft** has joined the staff of the Iowa State College Statistical Laboratory as Associate Professor—July 1, 1949. **Gobind Ram Seth**, on a four months' leave of absence from the Statistical Laboratory at Ames, flew early in July to visit the statistical institutions in Sweden and England, before returning to Delhi, India, where he will be teaching. **Oscar T. Kempthorne** was married in Vancouver, British Columbia, Canada, on June 10, 1949, to Miss Valda M. Scales of Coogee, New South Wales, Australia. Professor and Mrs. Kempthorne will be at home at 127 Stanton, Ames, Iowa, sometime in July.

B I O M E T R I C S

**The Biometrics Section of the
American Statistical Association**

TABLE OF CONTENTS

The Choice of a Response Metameter in Bio-Assay .D. J. FINNEY	261
The Validity and Meaning of the Results of Biological Assays N. K. JERNE AND ERIC C. WOOD	273
A Biological Assay of TuberculinsR. A. FISHER	300
On a One-Dimensional Diffusion Method of Assaying Antibiotic Substances and Its Fundamental Formulas MOTOSABURO MASUYAMA	317
Routine Computation of Biological Assays Involving a Quantitative Response M. J. R. HEALY	330
Queries	335
The Biometric Society	341
News and Notes	345

Material for *Biometrics* should be addressed to the Chairman of the Editorial Board, Institute of Statistics, North Carolina State College, Raleigh, N. C.; and material for Queries should go to "Queries", Statistical Laboratory, Iowa State College, Ames, Iowa, or to any member of the committee.

As the consideration of articles for publication in *Biometrics* is dependent upon referees, it is necessary for us to have *three* copies of each paper.

We are taking this opportunity to request that in the future all manuscripts be submitted in triplicate.

GERTRUDE M. COX, *Editor*

Officers American Statistical Association

President: Simon Kuznets; President-Elect: S. S. Wilks; Vice-Presidents: L. S. Kellogg, H. A. Freeman, D. S. Brady; Secretary-Treasurer: M. M. Flood; Directors: C. H. Goulden, Isador Lubin, L. J. Reed, F. F. Stephan, W. L. Thorp, L. L. Thurstone.

Officers Biometrics Section

Chairman: Margaret Merrell; Secretary: Max A. Woodbury; Section Committee: Joseph Berkson, Harold F. Dorn, Alton S. Householder, Jerzy Neyman and Joseph Zubin.

Editorial Board *Biometrics*

Chairman: Gertrude M. Cox; Members: C. I. Bliss, D. J. Finney, H. C. Fryer, O. Kempthorne, A. M. Mood, Horace Norton, G. W. Snedecor and Jane Worcester.

Membership dues in the American Statistical Association are \$3.00 a year, of which \$1.00 is for a year's subscription to the quarterly *Journal*, \$.75 is for a year's subscription to *The American Statistician*. Members of the Association may subscribe to *Biometrics* at a special rate of \$2.00. Dues for associate members of the Biometrics Section are \$3.00 a year, of which \$2.00 is for a year's subscription to *Biometrics*. Single copies of *Biometrics* are \$1.00 each and annual subscriptions are \$3.00. Subscriptions and applications for membership should be sent to the American Statistical Association, 1603 K Street, N.W., Washington 6, D. C.

Entered as second-class matter, May 25, 1945, at the post office at Washington, D.C., under the Act of March 3, 1879. *Biometrics* is published four times a year—in March, June, September and December—by the American Statistical Association for its Biometrics Section. Editorial Office: 1603 K Street, N.W., Washington 6, D.C.

THE CHOICE OF A RESPONSE METAMETER IN BIO-ASSAY

D. J. FINNEY

*Lecturer in the Design and Analysis of Scientific Experiment,
University of Oxford*

INTRODUCTION

UNLESS THERE ARE theoretical objections or empirical indications to the contrary, the statistician usually assumes that quantitative data which are put before him are normally distributed. If he is obliged to abandon the hypothesis of normality, he may choose some alternative specification of the distribution, or he may seek a transformation of the data (1) in terms of which the distribution is normal. The alternatives are closely related, for adoption of a transformation implies some assumption about the form of the original distribution and is an analytical convenience rather than an essentially different approach. Unless the data are very extensive, they are unlikely to discriminate satisfactorily between two or more equally plausible normalizing transformations (such as the square root and the logarithm); it is therefore natural to inquire how far conclusions drawn from a statistical analysis may be affected by the choice that is made. If the distribution is really normal in terms of one transformation, it cannot be normal for another unless the second is itself a *linear* transformation of the first, and in general numerical results obtained by the use of different transformations will not be identical.

Similar inquiries might be made about other common assumptions, such as those of linearity and homoscedasticity of regressions, often introduced by the statistician in order to form a mathematical model (on the basis of which the data may be subjected to statistical analysis) even though no biological theory demands that particular model. Few would deny that these may represent reasonable approximations to reality, none would assert their exact and absolute truth. Must the fact that different statisticians would not always agree on the choice of a normalizing or linearizing transformation for a particular body of data be regarded as a flaw in the vaunted objectivity of statistical analysis?

ASSUMPTIONS IN BIO-ASSAY

During the discussions that led to another paper (6), the relevance of this question to biological assay became apparent. The statistical analysis of the commonest type of dilution assay, that based on a linear regression of response on the logarithm of dose,* involves assumptions that, in terms of the dose and response metameters adopted for the analysis:

- (a) The distribution of responses for any fixed dose is normal.
- (b) The variance of responses for a fixed dose is independent of the dose.
- (c) The mean of the response distribution is linearly related to the logarithm of the dose, each preparation tested having its own line.
- (d) The line relating mean response and log dose for any test preparation is parallel to that for the standard.

These have been discussed in detail in (6), where they are listed as B5, B6, B4, A3 respectively. The first three relate to the mathematical model implicit in the form of statistical analysis usually made. The fourth is essential to the logic of a dilution assay, for which the test preparation must behave as though it were a dilution of the standard preparation: unless the regression lines (or curves, if the regressions are not linear) are parallel, no assay of the test preparation in terms of the standard is conceivable (3,4). None of (a), (b), (c) is ever known to be true *a priori*, though experience of a particular type of assay may establish a strong presumption that they are sufficiently near the truth for practical purposes. Nevertheless, some check on the basic assumptions should be available from the internal evidence of an assay. Serious disagreement between the data and (d) must be regarded as evidence that the assay is fundamentally invalid, whereas significant deviation from (a), (b), or (c) may be only an indication of invalidity of the particular statistical analysis adopted and may perhaps be remedied by use of other metameter transformations. Choice between the alternative explanations and remedies cannot be made entirely on the evidence of one assay, and must to some extent depend upon the experimenter's (or the statistician's) experience of previous similar assays.

DATA FOR STUDY

Dr. Jerne pointed out to the writer that his analysis of a prolactin assay, used as an illustrative example elsewhere (5), might be misleading.

*Valid assay systems can be constructed on other foundations, (4), and even violently different schemes may sometimes be required (9), the system described here is of wide applicability and provides adequate illustration of the theme

The final statement of a potency estimate and fiducial limits was based not only on the data but also on a series of assumptions not there mentioned, of which (a), (b) and (c) above were the chief, and in fact these assumptions were of doubtful validity in the example. A charge that assumptions affecting the conclusions have not been explicitly stated might be levelled against most statistical analyses which make use of normal or other standard distributions, for these are generally considered too obvious to need statement on every occasion. Criticism of the analysis of the prolactin assay, however, was undoubtedly justifiable, as inspection of Table 1 will confirm: the two very high responses suggest a positive skewness, and a variance ratio test shows the variance to be significantly higher for the high dose than for the low. Nevertheless, the analysis had not been performed quite unthinkingly. No adequate test of normality can be made on the small number of observations used for an assay, but general statistical theory teaches that means of several independent observations, under very wide conditions, have distributions more nearly normal than the distributions of individuals: in forming an estimate of relative potency, it is the means for different dose-groups rather than individual responses that are important. Again, experience has indicated that modification of the statistical analysis of an assay in order to allow for the inconstancy of the variance of responses adds much to the labour but does not make any appreciable difference to the conclusions unless the variation in variance is very great.

TABLE 1
DATA FOR AN ASSAY OF PROLACTIN

The responses are the crop-gland weights of pigeons (in 0.1g.)						
Dose of standard preparation (i.u.)			Dose of test preparation (mg.)			
1.25	2.50	5.00	0.125	0.250	0.500	
38	53	85	28	48	60	
39	102	144	65	47	130	
48	81	54	35	54	83	
62	75	85	36	74	60	
Totals.	187	311	368	164	223	333

As a second example, for which the skewness is less apparent, but which also shows indications of heteroscedasticity, data from an assay of

testosterone propionate by Pugsley (8) will be used. The design of the assay was similar, and the results are shown in Table 2.

TABLE 2
DATA FOR AN ASSAY OF TESTOSTERONE PROPIONATE

The responses are measures of comb-growth of capons (in mm.)						
Dose of standard preparation (γ)			Dose of test preparation (γ)			
20	40	80	20	40	80	
6	12	19	6	12	16	
6	11	14	6	11	18	
5	12	14	6	12	19	
6	10	15	7	12	16	
7	7	14	4	10	15	
Totals:	30	52	76	29	57	84

As a study of the extent to which the conclusions from such assays might be altered by the use of a response metameter different from that directly measured, each set of data was analysed for a series of metameters

$$y^* = y^i$$

where i was given a series of values between $+3$ and -3^1 . For any value of i , an estimate of variance within doses (with 18 degrees of freedom and 24 degrees of freedom respectively, for the two assays) was obtained, calculations were made of various quantities to be used in tests of deviations from the basic assumptions (a)-(d), and finally an estimate of relative potency, with its lower and upper five per cent fiducial limits, was formed. The method of analysis has been fully described, for $i = 1$, in Section 2 of (5). The transformation $i = 0$ is equivalent to $y^* = \log y$, and, in this paper, $i = 0$ will always refer to the logarithmic transformation.

VALIDITY TESTS

For neither assay are the data adequate to provide tests of deviations from normality. The transformations tried will greatly intensify any

¹The adjustments for variation in initial body-weight discussed in (5) were not used for the prolactin assay: they would have increased the arithmetical labour without adding useful information.

positive skewness when $i > 1.0$, and include extremes of skewness more marked than any that would ordinarily be encountered in any practical consideration of alternative response metameters. For $i < 0$, a negative skewness appears and becomes steadily more pronounced as i is decreased. Thus the range of metameters studied includes a region in which no skewness is apparent, and that is about the most that may be expected as a criterion of normality on so few data.

TABLE 3
SUMMARY OF VALIDITY TESTS
(for explanation, see text)

i	Assay of prolactin				Assay of testosterone propionate			
	z	Values of t for			z	Values of t for		
		Prepara- tions	Linearity	Parallel- ism		Prepara- tions	Linearity	Parallel- ism
3.0	2.38	0.80	0.54	-0.41	2.91	1.33	2.32	1.34
2.5	2.01	0.88	0.44	-0.37	2.37	1.40	2.07	1.37
2.0	1.64	0.99	0.30	-0.32	1.83	1.46	1.64	1.38
1.5	1.28	1.11	0.12	-0.24	1.29	1.48	0.98	1.36
1.0	0.93	1.24	-0.10	-0.12	0.74	1.42	0.08	1.31
0.5	0.58	1.38	-0.35	0.03	0.18	1.26	-0.94	1.21
0.0	0.24	1.50	-0.63	0.23	-0.38	0.99	-1.59	1.08
-0.5	-0.09	1.60	-0.91	0.46	-0.95	0.67	-2.55	0.95
-1.0	-0.41	1.67	-1.15	0.70	-1.53	0.36	-2.86	0.85
-1.5	-0.73	1.69	-1.32	0.93	-2.11	0.10	-2.90	0.80
-2.0	-1.05	1.69	-1.47	1.12	-2.70	-0.11	-2.78	0.77
-2.5	-1.37	1.67	-1.54	1.28	-3.29	-0.27	-2.57	0.78
-3.0	-1.70	1.64	-1.56	1.41	-3.88	-0.39	-2.34	0.79
10%	0.56	1.73			0.48	1.71		
5%	0.73	2.10			0.62	2.06		
1%	1.07	2.88			0.90	2.80		

Table 3 summarizes validity tests for the two assays. As a test of the independence of variance and response, the error mean square for the highest doses of both preparations has been compared with the error mean square for the lowest doses in terms of Fisher's z , with (6,6) and (8,8) degrees of freedom respectively for the two assays. Values of z for 10%, 5%, and 1% probability levels, on the null hypothesis that the true variances at the extremes of dose are equal, are shown at the bottom of the columns. For any value of i outside the range $0.7 > i > -1.5$ for the prolactin assay, or $0.9 > i > -0.2$ for the testosterone, the evidence against homoscedasticity would be judged significant, and for values of i very little more extreme the evidence is overwhelming. The further calculations for an assay, in their usual

form, are strictly valid only if the regression of response on dose is homoscedastic, but experience suggests that the difference between extremes of variance has to be very great before neglect of it has appreciable effect on the estimate of potency and its fiducial limits. The validity test just described is more sensitive to changes of metameter of the type under discussion than the other tests summarized in Table 3, and may often be unduly sensitive. In well-planned symmetrical assays of ordinary size, modification of the statistical analysis on account of heteroscedasticity is seldom needed unless the evidence against a constant variance is significant at least at the 1% level.

In order to avoid troubles arising from observations outside the range of linearity of the response curve, corresponding doses of the two preparations should be chosen so as to give responses as nearly equal as existing information on potency will allow. Failure to achieve equality is not itself an indication of invalidity, but is a warning that non-linearity may have serious consequences. Both assays under discussion are symmetrical, and a significance test of the difference between mean responses to the two preparations is therefore relevant to this point. Table 3 shows values of t for such a test, based on pooled variance estimates from all doses, under the heading "Preparations". For neither assay does t attain significance, even at extreme values of i , as is seen by reference to the values for 10%, 5%, and 1% probability at the bottom of the table.

The routine analysis of assays of this pattern makes use of measures of deviation from linearity (the mean coefficient of a quadratic term in the regression equation) and from parallelism (the difference between regression coefficients for the two preparations); values of t for these are also shown in Table 3, and may also be compared with the entries in the last three lines of the table. For the first assay, neither test shows invalidity at any value of i . For the second, there are strong indications of non-linearity except when $2.5 > i > -0.1$. Faced with such results, the statistician would see no reason to doubt the fundamental validity of either assay, but he would probably feel bound to regard his form of analysis for the second assay as invalid if it employed a value of i outside this range. It is of interest to note that these validity criteria do not always decrease or increase steadily over the whole range of i studied.

ESTIMATES OF POTENCY

The potency of the test preparation relative to the standard, together with its 95 per cent fiducial limits, were calculated on the hypothesis that (a)–(d) were true. The limits were obtained by use of Fieller's

theorem on the fiducial limits of a ratio (and not by the common approximation from a variance of the log potency), in which the quantity

$$n - \frac{t^2 V(b)}{g}$$

where b is the regression coefficient of y^* on log dose, plays an important part (2,4). A value of g that is little less than unity implies that b only just attains a magnitude significantly different from zero, and the contribution of errors in b to the errors of the potency estimate is therefore large; a small value of g indicates a relatively precise estimation of b , errors in which therefore make little contribution to the error of the potency estimate. In Table 4, g is tabulated as well as the potency estimates. For the prolactin assay, g is always larger than 0.2, and reaches a minimum near $i = -1.0$. For the testosterone assay, g is usually small enough to be negligible, and has a minimum near $i = 0.5$; with a large negative value of i , however, and probably also with a positive value larger than those studied, it becomes large.

TABLE 4
SUMMARY OF POTENCY ESTIMATES
(for explanation, see text)

i	Assay of prolactin			Assay of testosterone propionate		
	g	Potency	5% limits (i.u. per mg.)	g	Potency	5% limits (g. per g.)
3.0	0.71	6.96	0.37-22.5	0.050	1.18	0.91-1.54
2.5	0.60	6.92	0.85-18.9	0.038	1.16	0.93-1.47
2.0	0.49	6.88	1.39-16.4	0.030	1.15	0.94-1.41
1.5	0.40	6.84	1.91-14.6	0.023	1.13	0.95-1.36
1.0	0.33	6.80	2.36-13.3	0.020	1.12	0.95-1.32
0.5	0.28	6.76	2.72-12.4	0.019	1.10	0.94-1.29
0.0	0.24	6.71	2.97-11.8	0.020	1.08	0.92-1.27
-0.5	0.22	6.66	3.09-11.4	0.025	1.06	0.89-1.27
-1.0	0.22	6.58	3.08-11.2	0.034	1.04	0.84-1.28
-1.5	0.22	6.53	3.01-11.1	0.048	1.01	0.78-1.31
-2.0	0.24	6.42	2.81-11.2	0.071	0.98	0.72-1.34
-2.5	0.26	6.32	2.55-11.3	0.104	0.95	0.65-1.39
-3.0	0.30	6.19	2.22-11.5	0.148	0.92	0.56-1.46

The object of the assay is the estimation of the potency of the test preparation and the assignment of fiducial limits to the true potency. Figures 1 and 2 illustrate the manner in which these results are in-

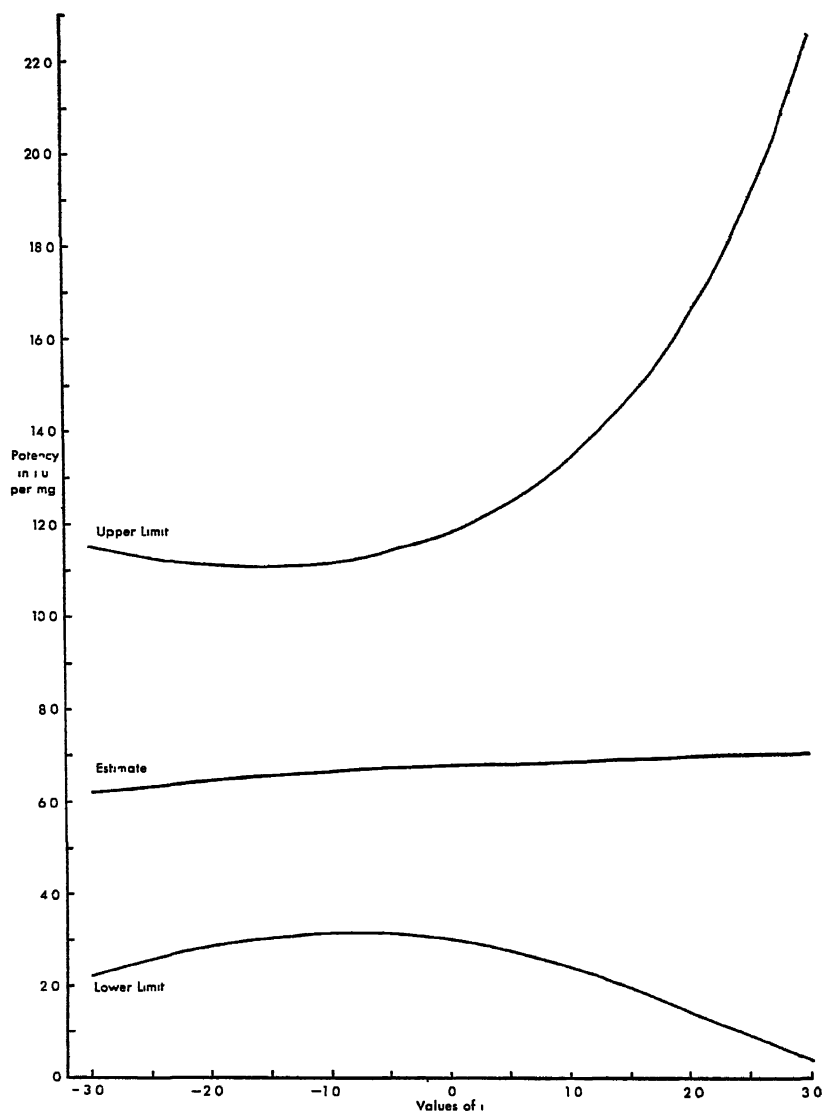


FIGURE 1.
ESTIMATES OF POTENCY AND FIDUCIAL LIMITS FOR PROLACTIN ASSAY.

fluenced by the choice of i . The potency estimate is not altered to any great extent by small changes in i . If shown the data from either of these assays, it is unlikely that any statistician would contemplate the use of a response metameter of the form y^* with $i > 2$ or $i < 1$, and over this extreme range the potency estimate changes only by 5%–10%.

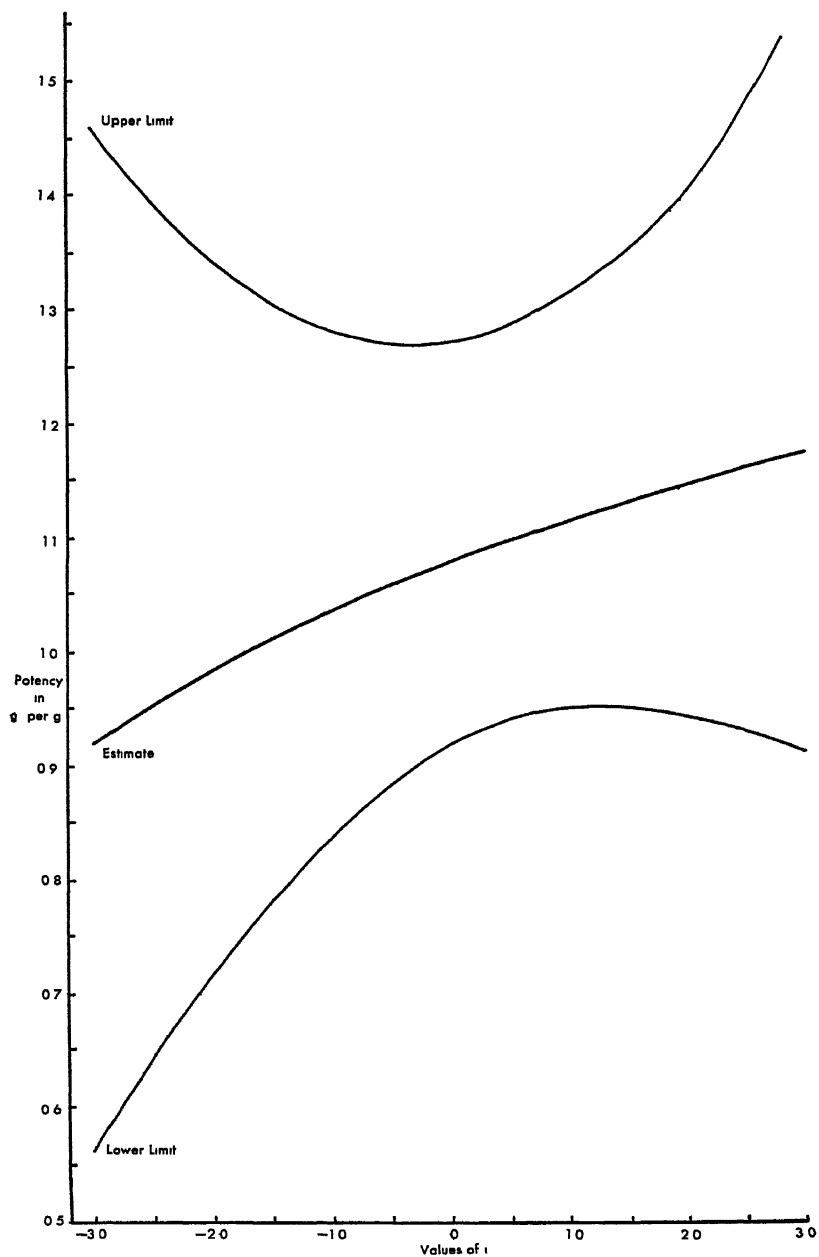


FIGURE 2.
ESTIMATES OF POTENCY AND FIDUCIAL LIMITS FOR
TESTOSTERONE PROPIONATE ASSAY.

In the absence of prior evidence to the contrary, his choice would almost certainly be made between $i = 1$ and $i = 0$, so that his subjective judgement would affect the estimate only by 2%-4%.

The fiducial limits are more sensitive to metameter changes. For the prolactin assay, the fiducial range is narrowest at about $i = -1.0$, and is appreciably widened (primarily because of the increase in g) if $i > 0.5$ or $i < -2.0$. For the testosterone assay, the range is narrowest at about $i = 0.0$, and begins to widen seriously for $i > 1.5$ or $i < -1.0$. There is, of course, no reason to regard narrowness of fiducial range as a criterion for the right choice of a response metameter, especially as the method of calculating the limits assumes the truth of conditions (a)-(d). If there were an exact correspondence between the mathematical model and the experimental data, exact probability statements could be made about the fiducial limits, but that is an ideal impossible of achievement. In practice, all that is required of fiducial limits is that, when computed by standard rules, they shall give an indication of a range within which the true value almost certainly lies. The user of a biological assay will wish to base some course of action on its results, but he would be unwise to base critical decisions on whether a particular value for the potency is just within or just beyond the calculated limits. For most of his questions, the limits will give a clear answer, but cases of doubt should be resolved by further experimentation and not by undue reliance upon the perfect truth of an abstract model. The main conclusion to be drawn from Table 4 and Figs. 1 and 2 is that, over quite a wide range of values of i , neither the potency estimate nor its fiducial limits (as calculated by the standard procedure) are affected by a change of metameter to an extent that would seriously affect decisions which had to be based upon the assay results.

CONCLUSIONS

It is not suggested that calculations of this kind should be undertaken as part of the statistical analysis of every biological assay. The investigation here reported was undertaken as a corollary to the work of Jerne and Wood (6), and as a warning to users of bio-assay against uncritical acceptance of a standard pattern of computation without thought of the assumptions involved. The theoretical implications of a particular choice of metameter must not be forgotten, however little alternative choices would alter the inferences made from the data.

The metameters tried for the prolactin and testosterone assays belong to a very restricted class, yet extensive calculations were needed for the construction of Tables 3 and 4. A larger set of possibilities would be comprised within the formulation

$$y^* = (y - y_0)^i$$

where both i and y_0 may be chosen so that the transformed data shall satisfy the basic assumptions of the assay and its analysis (7); another alternative would be a metameter that recognized the existence of both an upper and lower limit to possible values of y .

Tables 3 and 4 suggest that, important as are (a), (b), (c) to the logic of the statistical analysis, the practical conclusions from an assay will seldom be seriously affected even by violent changes in the response metameter: the transformations

$$y^* = y^2$$

and

$$y^* = y^{-1}$$

are sharply contrasted, yet in two examples they produce substantially the same conclusions on potency.

As might be expected, the z -test for heteroscedasticity is very sensitive to changes in the metameter, which may support Fieller's view that a transformation for equalizing variances is of prime importance. On the other hand, application of the z -criterion to the testosterone assay would lead to rejection of the analyses based on certain metameters when in fact the results of such analyses in respect of the potency estimate are quite satisfactory. The other criteria proved surprisingly insensitive to changes in i for metameters of the type under discussion. The response relationship for the second assay would indeed have been regarded as significantly non-linear except for a restricted range of i , a range that does not agree very well with the indications of statistical validity from other sources, but even the most extreme transformations failed to disturb the parallelism criterion. Other types of metameter might disturb the other validity criteria more severely, but evidently tests of parallelism and linearity, in assays of ordinary size, will often fail to disclose invalidity because of their low sensitivity.

The present analyses may be regarded as empirical evidence in support of the contention of Jerne and Wood (6) that metameters should be chosen on the basis of past experience. The confusion that follows from any attempt to determine the ideal metameters for a single, probably rather small, series of experimental measurements is evident from examination of Table 3 and 4, Figs. 1 and 2. Before an assay is performed, there should be strong reasons for believing in its fundamental validity and knowledge of suitable dose and response metameters to use in the analysis. The validity tests should then be regarded as a confirmation that no abnormal behaviour of the subjects or other disturbance from unknown causes upset either the fundamental

or the statistical validity of the assay, and not as a demonstration of the validity of a particular set of assumptions peculiar to one assay. Two different transformations will never lead to *exactly* the same numerical values for the estimate of potency and the fiducial limits, and, in the absence of information on which of the two is correct, the problem of deciding which is the *best* set to choose as summarizing the data, on the internal evidence of one assay, seems insoluble: if the standpoint just recommended be adopted, there are good grounds for believing that, in spite of the fact that two statisticians faced with the same data would not necessarily use the same metameters, the problem is of importance only to the philosopher, and that the decisions of the experimenter will not be affected by it to any marked extent.

SUMMARY

The effect of choosing various alternative functions of an observed response as a metameter in the analysis of biological assays is discussed. The argument is illustrated by multiple analyses of two assays, and conclusions are drawn relating to the choice of a metameter and the interpretation of validity tests in general.

I am indebted to Dr. N. K. Jerne for the suggestion that multiple statistical analyses of the results of an assay, using a series of different metameters, might be instructive. I wish to record my gratitude to him and to Dr. E. C. Wood for valuable exchanges of ideas over a long period and for helpful criticism of a draft of this paper.

REFERENCES

1. Bartlett, M. S. (1947). "The use of transformations", *Biometrics*, **3**, 39-52.
2. Fieller, E. C. (1939). "The biological standardization of insulin", *Journal of the Royal Statistical Society, Supplement*, **1**, 1-64.
3. Fieller, E. C. (1947). "Some remarks on the statistical background in bio-assay", *Analyst*, **72**, 37-43.
4. Finney, D. J. (1947). "The principles of biological assay", *Journal of the Royal Statistical Society, Supplement*, **9**, 46-91.
5. Finney, D. J. (1948). "The adjustment of biological assay results for variation in concomitant observations", *Journal of Hygiene*, **45**, 397-406.
6. Jerne, N. K., and Wood, E. C. (1949). "The validity and meaning of the results of biological assays", *Biometrics*, **5**, 4.
7. Kapteyn, J. C. (1903). "*Skew Frequency Curves in Biology and Statistics*". Groningen: P. Noordhoff.
8. Pugsley, L. I. (1946). "The application of the principles of statistical analysis to the biological assay of hormones", *Endocrinology*, **39**, 161-176.
9. Thompson, W. R. (1948). "On the use of parallel or non-parallel systems of transformed curves in bio-assay: Illustrations in the quantitative complement-fixation test", *Biometrics*, **4**, 197-210.

THE VALIDITY AND MEANING OF THE RESULTS OF BIOLOGICAL ASSAYS

N. K. JERNE

*Dept. of Biological Standardisation,
State Serum Institute, Copenhagen*

and

ERIC C. WOOD

*Virol Ltd., Ealing, London, W.5.**

INTRODUCTION

IN MAY 1948, one of us (N. K. J.) wrote to Mr. D. J. Finney making certain comments on his paper [13] on the adjustment of biological assay results for variation in concomitant observations. Mr. Finney, after replying to this letter, discussed the points raised with the other of us (E. C. W.), and a tripartite correspondence ensued in the course of which almost every link in the long chain of reasoning and inference from the performance of a biological assay to the statement of its results was critically examined. It seemed to us, whose own thinking on the subject had been much clarified, that the main conclusions at which we had arrived might usefully be brought together and placed on record for the benefit of those who might not have had occasion to give so much thought to the theoretical background of biological assays.

This background is surprisingly complex and varied. Behind the experimental design of an assay, and still more behind the final statement about the result and its fiducial limits, there lies a whole host of assumptions and implications. Some are chemical or biochemical, about the nature of the Test Preparation (T.P.) whose potency is to be assayed and that of the Standard Preparation (S.P.) with which it is to be compared. Some are biological or pharmacological, about the response evoked in the experimental animals by the stimulus of the doses administered to them. Some are mathematical and in particular statistical, about the computational processes by means of which the numerical results are evaluated. Some, even, are philosophical—if not metaphysical!—about the relations between the theoretical abstractions of

* Address altered since this paper was written to the County Laboratories, Redwell St., Norwich, Norfolk.

pure mathematics and the realities of bio-assay. Not all of these assumptions appear to have been explicitly stated before; only in some instances has their validity been discussed [4, 8, 16] and the consequences examined that would follow if in a particular assay one or more were invalid.

The most fundamental assumptions of all are three which it is not proposed to do more than state—

1. That the fundamental concepts of biology and chemistry as applied to the methodology of bio-assay are correct.
2. That the fundamental concepts of the theory of statistics as applied to the design and computation of bio-assays are correct.
3. That the application of the principles of statistics to data from biological experiments is legitimate and useful.

Discussion of these three tenets is outside the present scope; we shall take them as our starting-point and shall address ourselves to those for whom they are almost axiomatic.

Most of the symbolism, much of the nomenclature, and some of the phraseology of this paper is taken directly from the writings of Finney, particularly his address to the Royal Statistical Society in 1947 [12]. Moreover, his part in the correspondence referred to above has been decisive in preventing errors and fallacies from creeping in, and several of the points made below are his in their origin. While the presentation and style of this paper is our own, Finney's ideas permeate it to such an extent that he should be considered its godfather if not actually one of the parents. Valuable suggestions were also made by Dr. G. Rasch, and we have benefited much from discussions with him. Others, of course, have written on various aspects of bio-assay, and where we have consciously drawn upon these publications due acknowledgement has been made. If we have ignorantly overlooked some other anticipation of our remarks we offer our apologies in advance.

In what follows, it is supposed that a dose z (z_i if of the S.P., z_i if of the T.P.) is administered to the experimental animal or *test subject* and that this stimulus evokes some quantitatively measurable and continuously variable *response* u (or u_i or u_i). Quantal or 'all-or-none' assays are not considered, though much of the discussion should be quite applicable, *mutatis mutandis*, to such assays. If a number of test subjects be given the same dose of the same Preparation under constant experimental conditions (such a group will be called a *dose-group*) the response will vary from subject to subject because of experimental and sampling errors (in bio-assays, the latter will usually be so much larger than the former as to constitute the major cause of the

variation), and the symbol U denotes the expected or 'ideal' response of the population of test subjects to the dose z . The set of responses from any one dose-group will be called an *array*. An assay (or bio-assay) is defined as an experiment designed for the purpose of determining the 'potency' (exactly what this means is described below) of a Test Preparation relatively to some Standard Preparation from the body of experimental data collected, by utilising the fact that the two Preparations have the same qualitative effect on the test subjects.

We propose from now on to distinguish between the *fundamental validity* of the assay and the *statistical validity* of the computations which follow it. In Section A, below, the assumptions essential to fundamental validity are discussed; for example, the need for a proper design of the assay. If any one of these assumptions is untrue for a particular assay, the data obtained cannot lead to a correct answer, no matter what arithmetical processes are applied to them. If, however, one of the essential assumptions for statistical validity (see Section B) is untrue, then it is the method of computation that must be amended; the assay data are satisfactory but the means adopted of extracting from them the information sought are inappropriate. For the final statement of the relative potency and its fiducial limits to be 'valid', *tout court*, both fundamental and statistical validity are essential.

SECTION A—THE FUNDAMENTAL VALIDITY OF THE ASSAY
PART 1. THE ESSENTIAL ASSUMPTIONS

The assumptions essential to fundamental validity have been dealt with by several authors [2, 9, 10, 12, 16, 17, 19, 20, 22, 23]. Nevertheless, certain points have emerged from the correspondence mentioned above which suggest that it would be worth while to re-state these assumptions and to add a few comments.

A1. The differences between the several arrays in an assay are wholly caused either by differences in dosage or by random sampling; in other words, had the same dose been given to every test subject the arrays would have been random samples from the same population.

This assumption may be modified if certain factors affecting the response are known and recorded in such a way that their influence can be taken into account (e.g., by an analysis of covariance), in which event such factors may not be randomised.

A2. The expected response U is a function of the dose z , so that

$$U = F(z), \quad (a.1)$$

where U is 'a single-valued strictly monotonic function of z , at least over the range of doses to be used' [12] [Finney's Condition I.]

This condition is not essential, but is rather a restriction of convenience. The possibility that a certain value of the response might be given by either of two different doses implies obvious risks of fallacious conclusions.

If the S.P. and the T.P. both satisfy this condition for a certain range of responses, so that

$$U_s = F_s(z_s) \quad (\text{a.2})$$

and

$$U_t = F_t(z_t), \quad (\text{a.3})$$

then for any selected value of U within the range in question, doses z_s and z_t can be found such that the expected response to each is the same, U . At this level of response it can then be said that the T.P. has z_s/z_t times the *potency* of the S.P., and this is within its limitations an unambiguous definition of potency. But because no assumption has yet been made about the relation between functions $F_s(\)$ and $F_t(\)$, no statement can be made about the potency at any other level of response.

A3. If the substance in the S.P. responsible for evoking the characteristic response from the test subjects is called the *effective constituent*, then the response evoked by the T.P. is due solely to the presence in it of the same effective constituent without modification by any other substance, 'so that the less potent preparation behaves as though it were a dilution of the other in a completely inert diluent' (Finney's Condition II).

From this it follows necessarily that, apart from the effects of sampling and experimental errors, the estimate of potency obtained by a valid assay procedure is independent of the experimental conditions or environment, the measurement chosen as the response, and the species or variety of test subject used. It is worth making this statement explicitly, because this assumption is made whenever an estimate of relative potency obtained from rats or guinea-pigs is applied to the administration of the preparation assayed to cattle or human beings.

The ratio $\rho = z_s/z_t$ must, if this 'hypothesis of similarity' [23] be true, be independent of U , for it represents the relative amounts of the effective constituent in equal doses' [12]. Thus $z_s = \rho.z_t$ always, and equation (a.3) can be re-written

$$U_t = F_t(\rho.z), \quad (\text{a.4})$$

where ρ is the potency of the T.P. relatively to that of the S.P. at *any* level of response; this is the only definition of relative potency that

would normally be regarded by the bio-assayist as satisfactory. As Finney says [12], 'If the data cannot be adequately described by the same form of $F(\)$ for both preparations, the basic assumption that only the same effective constituents were concerned in both must be false. . . . The assay is therefore invalid'—and it might be added that the whole idea of assaying that particular T.P. against that particular S.P. becomes absurd.

Similarly, equation (a.4) must be true for any kind of test subject and any measurement chosen as response; the form of the function $F(z)$ may well change as the subject or the measurement is changed, but ρ must be invariant.

The 'effective constituent' can be a mixture of two or more distinct chemical compounds, provided they are in fixed proportion in the S.P. and T.P.

It may be worth while, as showing that this discussion is not so trite as to be superfluous, to quote some assay procedures in which assumption A.3, for one reason or another, does not hold *in toto*. When a preparation containing vitamin D₃ is assayed against a S.P. containing vitamin D₂ using rats, apparently valid results are obtained, but the 'potency' of the T.P. is much less than if chicks are the test subjects. Here the less potent preparation certainly *behaves* like a dilution of the other in an inert diluent, but in fact it is not; the effective constituents are not the same. Again, remembering that the nutritional requirements of various species differ, it might easily happen that the 'inert' diluent might be so for one kind of test subject but not for another. A variant might occur where the action of a food when given as a T.P. is indirect, stimulating the growth in the digestive tract of bacteria which in growing synthesise some nutrient utilised by the host. In other species not harbouring such bacteria, results might be different. An assay of the vitamin B₁ content of *live* yeast by feeding tests, using a reference sample of *killed* yeast of known vitamin content as S.P., might yield curious results. If given to a species whose gastric juice was so acid as to kill live yeast before it left the stomach, a valid estimate of the true vitamin B₁ content should be obtained; otherwise, the result might be quite erroneous, for yeast growing in the gastrointestinal tract absorbs vitamin B₁ from the foods undergoing digestion and thereby acts towards the host as a depletor of the vitamin. An assay in such circumstances might well show an apparently *negative* vitamin B₁ content!

An instance of current interest in which assumption A.3 does not hold is the assay of diphtheria and tetanus toxoids in commercial products containing aluminium hydroxide, using as S.P. a reference

sample of highly purified toxoid. The adsorbent $\text{Al}(\text{OH})_3$ in the T.P. is not inert but interacts strongly—in an unknown manner—with the 'effective constituent'. Even so, within a certain range of dosage, the less potent preparation behaves like a mere dilution of the other, but when a wider range is examined it is found that the corresponding range of responses is much wider for the Al-adsorbed toxoids than for the plain toxoids, i.e., the dose-response curves of the two preparations have different upper asymptotes and cannot be described by the same form of $F(z)$. The assay is thus invalid.*

This is an illustration of the fact that in practice one is usually aiming at estimating the relative potency of two preparations supposed to contain the same 'effective constituent' but differing somewhat in 'diluent'. It may well happen—see above—that the diluent is inert for some species but not for others. In such cases the distinction between the effective constituent and the diluent tends to disappear; if only one kind of test subject be used, one can never be sure whether the comparison of the S.P. with the T.P. is based on one effective constituent only or on two, in differing proportions. This can and does occur in assays of, e.g., hormones and sera.

By comparing the T.P. with the S.P. using several different kinds of test subjects, concordant results may inspire confidence in the essential similarity of the two preparations; the reverse is equally true. This clearly implies that a substantial biological research must precede any attempt to set up a bio-assay technique for routine use.

It should never be forgotten that bio-assay is not itself a basic science but an applied science, depending almost entirely upon biological research. When a bio-assay is conducted involving reactions that have been inadequately studied, the risk involved in making the assumptions discussed in this section is very great, and one's confidence in the results obtained should be correspondingly small.

PART 2. RULES OF CONDUCT BASED ON THESE ASSUMPTIONS

A1. In every assay there must always be many factors other than dose affecting the response. Some of these will be known and others unknown. Taking first the known factors, there are three ways of dealing with them.

First, there are some which it is both possible and convenient to hold constant, or substantially constant, for all the test subjects through-

*Thompson [21] has shown that useful conclusions can be drawn from certain types of assay in which a substance other than the 'effective constituent' exerts an effect on the response and our assumption A3 is thus quite untrue. We are not concerned with such assays, which are outside the scope of this paper.

out the assay period. Examples are afforded by the size of cage, the atmospheric temperature, the amount of handling by the attendants.

Secondly, there are others which are known to be important but cannot be held constant. Each such factor should be recorded for every test subject so that the information can be properly utilised in the final calculations, either by an analysis of covariance or by segregating that factor and its interactions in the analysis of variance. The experimental design, and the allotment of individual test subjects to the various dose-groups, must be such as will permit one of these two alternatives to be employed. Such factors as sex, litter, and body-weight might well come into this category in particular assays.

Thirdly, other factors known or assumed to be unimportant are best dealt with by randomising them, so that their effect on the response, though unknown, will not introduce bias into the result of the assay. Minor differences of age between test subjects, the order of dosing them, and so on, are factors which may properly be randomised.

The point ought to be made, however, that randomisation is not the best way of dealing with a factor that ought to be and could be included in the previous category of recorded factors. Randomization leaves the worker ignorant of the effect of the factor randomised upon the response, and this knowledge is valuable both in the assay itself and in planning future assays. Moreover, if the influence of the factor is material, randomisation implies an additional assumption—that the contribution of the factor to the variance of the transformed response is truly additive—and if this assumption is incorrect the reliability of the assay is decreased. An illustration is the use in an assay of animals some of whom are grouped closely around one initial body-wt. (say 600g.) and the rest around another (say 400g.). If body-weight affects the response, *random* distribution might result in a series of two-topped array distributions and ruin the assumption of Normality (B3 in Section 2).

Care must also be taken to ensure that where randomisation is attempted it is actually achieved. Departures from the 'strait and narrow path' of pure randomisation do manage from time to time to insinuate themselves most subtly. For instance, Emmens has pointed out [8] that to put one's hand into a cage of animals and take the first one caught for the first dose is *not* random selection unless the doses themselves are randomised, for presumably the animals that are caught last are the most alert and vigorous in eluding the pursuing hand and may differ in responsiveness to dosing from their more easily caught, perhaps because more weakly, associates. With orderly dosing techniques, all the most easily caught animals will be in the first dose-

group and all the most elusive in the last. Similarly it is usual in microbiological assays to dose successively all tubes in the first dose-group and then pass to the next; when this deliberate non-randomisation does not cause trouble it is only because the assay tubes can be made far more nearly replicas of each other than can ever be attained with the macro-biologist's test material.

Occasionally, evidence is obtained from the data of an assay that assumption A1 is not valid for that assay; this possibility, and the consequences that ensue, are discussed in Section B, Part 2.

A2. The assumption that *some* function connects the response with the dose, and that for each given value of the expected response within the range used there is one value and one only of the dose, can justifiably be taken for granted in the absence of evidence to the contrary. One can imagine circumstances in which the responses to two different doses of the same preparation were identical, but the point in practice would not cause any trouble. The question of the nature of the function is quite another matter; it is dealt with below.

A3. The Hypothesis of Similarity [23] between the S.P. and T.P. can be tested statistically by examining the identity in form of the functions connecting response with the dose of the two preparations, provided that the experimental design permits. The methods are well-known and will not be discussed in full here; in assays in which some function of the response is linearly related to the log. of the dose, the parallelism of the two regression lines is examined, while in assays in which the linear relationship is to the dose itself, the intersection of the two lines at the zero-dose level is the criterion. The greater the dose-range covered by the assay, the more sensitive is the test; but since too great a range of doses would involve risk of exceeding the limits over which the linear relation holds, a compromise is necessary. In theory, it does not matter whether the relation between the chosen function of the response and of the dose is linear or not; but the test for identity of form would be cumbersome with other than linear curves and as the calculation of the result and its fiducial limits would also be complicated, linearity becomes itself an assumption (see below).

Whatever statistical criterion may be applied to test the hypothesis that the two response curves are identical in form, it can only be 'disproved' or 'not disproved'—it can never be 'proved'. Moreover, the borderline between the two possible answers will depend on the degree of probability taken as 'significant'. The analyst must use his own judgment in this respect; he would be justified in accepting as valid an assay of a T.P. whose composition was known *a priori* to be very

similar to that of the S.P., particularly if many T.P.'s of the same kind had been validly assayed before by the same technique, where he should reject as invalid, or at least reserve judgment on, an assay giving exactly the same numerical value of 'validity criterion' but in which the T.P. was of an unfamiliar nature and the assay was novel in type. In other words, the statistical calculations present the evidence objectively and efficiently, but the decision is taken by the analyst in the light of his experience and judgment.

The statistical tests for validity may give misleading results unless the mean responses to the doses of T.P. and those to the doses of S.P. cover about the same range. If the S.P. responses all fall within the limits of linearity but the T.P. responses do not, non-linearity of the T.P. curve may cause the rejection of a good assay for apparent non-parallelism or non-linearity; worse still, a fundamentally invalid assay might fail to be rejected because of accidental agreement in slope of the two response curves over parts of their respective ranges that were not truly comparable. Occurrence of a significant difference between the mean responses to S.P. and T.P., though not an indication of invalidity in itself, is a warning that the other criteria of invalidity must be scrutinised with especial care.

In some instances, as when the T.P. and S.P. are known to be dilutions of the same pure compound in the same diluent, the truth of assumption A3 is self-evident. On the other hand, there are assays which are known not to be universally valid, and for which a statement of potency ought to be accompanied by the name of the species for which it is applicable. But in the majority of assays for which A3 is accepted, this is done simply because there is no evidence to the contrary. Therefore, whenever it is possible to carry out assays using different test subjects, responses or techniques, or to perform parallel analyses by chemical or physical methods, the additional evidence thus provided is most valuable.

SECTION B—THE STATISTICAL VALIDITY AND EFFICIENCY OF THE COMPUTATIONS

PART 1—THE ESSENTIAL ASSUMPTIONS

Section A deals with the matters on which the analyst must satisfy himself before he is prepared to make any statement whatever about the potency of the T.P. Equally complex considerations are involved when deciding the precise numerical values to be quoted as the 'best' estimate of potency and its fiducial limits. These will depend on several assumptions about the nature of the dose-response relationship and the computations employed. The assumptions do not, of course, affect the

fundamental validity of the assay; but *some* assumptions must be made before any computations can be performed at all, and those on which the calculations at present in use are most often based will now be enumerated.

B1. The doses both of the S.P. and T.P. have been measured sufficiently precisely for errors of measurement in the dose z to be negligible in comparison with the sampling and experimental errors in the response u .

This assumption is inherent in the usual formulae by which the regression equations are calculated. The computations could no doubt be modified for assays in which this assumption does not hold.

B2. There exists at least one function y of the response u , and one function x of the dose z , such that (a) to each value of u and z within the range of the assay there corresponds one and one only real value of the response metameter y and the dose metameter x respectively; (b) the transformation of u to y is independent of the transformation of z to x ; (c) the values of y and of x so obtained satisfy assumptions B3 to B6 inclusive.

Possible forms for the functions include, of course, $y = u$, $x = z$.

B3. The functions so defined are known and are of such a kind that y and x can be computed.

In theory, the functions instead of being known completely (e.g., $x = \log. z$, $y = u^2$) could be known in terms of a parameter or parameters to be estimated from the assay data ($x = \log. (n + z)$, $y = u^m$). The computations would thereby be extended considerably.

B4. The relation of Y , the expected or 'ideal' value of y , to x is exactly expressed by the linear equation

$$Y = a + b.x, \quad (\text{b.1})$$

(in which a and b are constants to be evaluated from the assay) over the range of the observations used in the calculations. (Assumption of Linearity.)

The relation between Y and x could be non-linear but precisely known. Once again, the calculations would be complicated considerably; Fieller [9] has shown how curvature can be allowed for.

B5. For any one value of x within the range of doses used the frequency distribution of y is exactly normal. (Assumption of Normality.)

Alternatively, the distribution might be not Normal but of other known mathematical specification—a possibility not likely to be of much practical importance when the response metameter is a continuous variate.

B6. The variance of y is independent of Y . (Assumption of Homoscedasticity.)

This assumption, though usual, is not essential; the variance could be assumed to depend in a known manner on Y . The consequences of doing so are discussed below.

B7. Provided assumptions B1 to B6 are true, the mathematical process actually applied to the data leads rigorously to exact and unique values for the 'best estimate of potency' and its 'fiducial limits'.*

This implies that the computations employed are those appropriate to the data of the particular assay being evaluated; that all available information (including, e.g., concomitant measurements) has been efficiently utilised; that arithmetical blunders have not been committed!

PART 2. RULES OF CONDUCT BASED ON THESE ASSUMPTIONS

B1. Very little comment is necessary on the obvious requirement that the method of dosing used must ensure that each and every test subject receives its intended dose with high precision and accuracy relatively to the measurement of the response. If vitamins, for example, are being given *per os*, it must be certain that the whole dose is delivered into the animal's mouth and swallowed without loss. There are a few assays in which the error of measuring the dose is unavoidably large; for example, when the dose is measured as numbers of bacteria injected, the magnitude of each dose is dependent upon a plate count of bacteria and is thus subject to considerable error. The magnitude of the error in x should then be estimated separately if possible and the computations modified accordingly.

B2, B3. The existence of, and the practicability of formulating, the equations for transforming the dose and response as actually measured into metameters amenable to standard computational processes may be taken for granted. It is true in theory that there may be no pair of transformations for which assumptions B3 to B6 are true simultaneously. In practice, there will often be many—perhaps an infinite number—of transformations for which the critical assumptions are satisfied sufficiently closely (see below). For example, if $y = u$, $x = \log z$, are found satisfactory, then there is no doubt that small variations such as $y = u^{1-\delta}$, $x = \log(z + \delta)$, where δ is a relatively small quantity, would give equal satisfaction, to name no other possibilities. There is no logical reason for preferring one of these satisfactory pairs of transformations to any other when dealing with *a single assay considered in isolation*; a transformation should be selected for which

*For the meaning of these terms, see Section C.

(first) there is no significant evidence that the assumptions B3 to B6 do not hold good, and (second) the subsequent calculations are as easy as possible. Too much must not be made of this second point; compared with the time taken for a full computation of fiducial limits, it is of no consequence if a few more minutes are spent transforming the measured response into, say, $\sqrt{(u^2 + 5)}$ instead of u itself, provided that some advantage is to be gained thereby.

If, of course, there are plausible reasons based on chemical or biological theory for assuming a particular relationship between the dose and the response, this may provide a basis for the selected transformation; but it must be emphasised that the transformations used in practice at the present stage of biological assay are almost always purely empirical, and the reasons for their selection pragmatic, at least in the first instance. Conclusions drawn from such single assays must therefore always be regarded as somewhat tentative.

The position is materially altered when the assay to be considered has a background of previous experience—when it is one of a series based on an identical methodology extending into the past and the future, or when there has been a prior research into the nature of the dose-response relationship under the conditions used in the assay. Some of the assumptions can then be tested far more rigorously than from the data of an isolated assay. The assumption of Linearity (B4) may not be significantly departed from in any one assay of a series; but if twenty consecutive assays deviate *in the same direction*, the pooled evidence effectively disproves the assumption for the series as a whole. Again, the variance of y may show a tendency to increase as x increases which is quite without significance in any one assay but because of its occurrence in every assay in the series demonstrates that assumption B6 is untrue. When, therefore, a number of similar assays has been performed, a transformation ought to be selected for which it has been shown by appropriate tests* that departures from each and every one of the assumptions made will be as often in one direction as in another, so that the results quoted for the assays will be as often too high as too low, and over the series of assays will be without any systematic bias.

B4. If there is no significant departure from the linear relationship for either S.P. or T.P., and provided that such deviations as do occur in a series of assays show no consistent trend in direction or type, the assumption of Linearity may be made. On the other hand, it is cer-

*Or for which the bio-assayist feels convinced, on the basis of a long experience, without making tests. . . . It is to be feared, however, that many workers (including the present writer!) tend to omit the tests on the slightest excuse—usually pleading lack of time.

tainly not justifiable to take it for granted without evidence, as would be the case if, for example, a 4-point assay design (two doses of S.P. and two of T.P.) were used without other evidence—this is discussed more fully below. Even when it has been established by exploration of the full dose-response curve that there is a range of doses over which the relation between the dose metameter and the response metameter is effectively linear, a watch must be kept on subsequent assays to ensure that over a period of time the inevitable changes in the sensitivity of the animal colony to the preparations do not lift or lower the mean responses as a whole on to a non-linear part of the dose-response curve. This means that there should be at least three dose-groups each of the S.P. and T.P. In assays in which one animal of each litter is placed in each dose-group, in order to obtain the well-known advantages of being able to segregate inter-litter variance, the litters may not all be big enough to allow of this being done for more than two doses of each preparation. Even so, some litters will contain more than this minimum number of animals, and these can be used to form small dose-groups additional to the assay proper; their responses may not be used in the final calculations, but the evidence they provide, and the sense of security thereby engendered, make the extra work well worth while. Alternatively, incomplete block designs such as have been developed for agricultural experiments can be employed. If a long series of routine assays of the same kind is continually being extended, it is very desirable to carry out at intervals a check on the full dose-response curve, using as many doses of the S.P., and over as wide a range, as possible, falling above and below the expected linear range as well as within it.

The 4-point design should be used only when there is either previous well-substantiated experimental evidence or *a priori* knowledge that the S.P. and T.P. are of the same nature, i.e., that the Hypothesis of Similarity (assumption A3) is justified; for in this design departure from this vital hypothesis cannot be distinguished from non-linearity, as has previously been mentioned. On the other hand, if all the conditions for validity were *known* to be fulfilled, this design would certainly make the most efficient use of the animals available, or putting this another way, it would enable a specified precision to be attained with a minimum number of animals. There are thus sometimes circumstances in which it is the design of choice; but its drawbacks, and the risks involved in using it without adequate safe-guards [11], must be emphasised.

With other designs, the Between Treatments component of variance has 4 or more degrees of freedom, so that it is possible after taking out the Linear Regression fraction and the 'S.P. v. T.P.' fraction to examine

separately the contributions made by departures from assumptions A3 and B4. (In assays in which x is the logarithm of the dose, these are usually referred to by the convenient terms of Unparallelism and Curvature respectively [8], and we shall use these terms here.) Significant curvature calls for further examination of the data to decide its nature, for the action to be taken depends largely on the answer to this question. Three principal cases can be distinguished.

(a) Both the S.P. and T.P. lines are curved, the curvature being in the same direction and of approximately the same magnitude in both—in assays in which ‘Mean Curvature’ and ‘Opposed Curvature’ can be examined statistically using orthogonal comparisons, this means that the first is significant and the second not. This suggests that the transformations used to obtain y and x are inappropriate; if some other can be found which will linearise the relationship between the dose and response metameters for the S.P. it will almost certainly do so for the T.P. as well. The assay should be quite valid, however; there is no suggestion that assumption A3 does not hold.

(b) Both lines are curved, but in opposite directions, or one is much more curved than the other, i.e., ‘Opposed Curvature’ is significant. This means that the assay as it stands is certainly invalid; either the Hypothesis of Similarity is not true, or it appears to be untrue because the range of doses for which the linear relationship holds has been exceeded. This can be checked by inspection and trial; it may be that the assay can be linearised by omitting the highest (or the lowest) dose-group on one or both preparations. If what is left is a 4-point assay, of course, it may be of doubtful value unless there is other evidence available as discussed above.

(c) Neither line is curved but the mean responses deviate from their respective regression lines in a random although significant manner, i.e., the ‘Between Dose-groups’ deviations are significantly greater than the ‘Within Dose-groups’ variance and yet are not accounted for by any kind of uniform curvature. This implies that proper randomisation has not been attained either in assigning subjects to doses or in the errors attaching to dosing, i.e., that assumption A1. does not hold. The moral is to alter the technique employed in subsequent assays; but in calculating the results of the present one, allowance may be made for the high residual variance between dose-groups (after removing the variance attributed to Preparations and Regression) by taking this, rather than the within-groups variance, as the square of the quantity to be used as the standard error in computing the fiducial limits of the assay. This is to be regarded as a device for making the best of a bad

job, not for permitting deliberate and persistent flouting of the need for randomisation.

Small departures from Linearity have very little effect on the estimate of potency, and if this with its fiducial limits is computed ignoring the non-linearity component of variance the result will be very close to the truth. But the experimenter must be sure that there is no systematic bias in his results as a whole due to a persistent trend in the curvature, and above all he must be satisfied that what he takes as non-linearity is not something far more serious—namely, invalidity of the Hypothesis of Similarity.

B5. The assumption of Normality is almost impossible to test from the data of any ordinary assay, because the replications available are far too few, and unless there is something very odd indeed about the responses as measured (and this would lead the experimenter to discard the assay in any event!) the assumption is taken for granted. The consequences of doing so when in fact there is a significantly non-Normal distribution have been discussed both in the general case [7, 17] and for the special circumstances of biological assay [10, 12]. It is reassuring to find that the result of the assay should be almost unaffected by quite large departures from Normality, because the mean slope and the mean difference in response between T.P. and S.P., the quantities from which the estimate of potency is derived, are means based on all the observations, and it is known that such quantities are much more nearly Normally distributed than are individual observations. But this is not true of the fiducial limits of the estimate of potency; they are functions not only of the two quantities just mentioned but also of others which are not means. One of them is Student's t , and Geary [18] has shown that the use of t in tests of significance may lead to quite erroneous conclusions if the relevant population is skew. Further investigation appears to be called for, first by the experimenter to examine the frequency distribution of such responses as are used in practical assay techniques, using as large a number of test subjects as possible in relatively few dose-groups, and secondly by the mathematician to discover how far non-Normality if it occurs in biological assays affects the fiducial limits calculated by the standard formulae assuming Normality.

B6. The assumption of Homoscedasticity is usually made in the absence of any significant evidence to the contrary. It must be admitted that the point is rarely examined by any more stringent tests than 'eyesight' inspection. Fieller [10] has given it as his opinion that this is the primary requirement to be satisfied by the response meta-

meter and that other requirements are secondary, but Finney [12] is not in complete agreement with him on this point. If the assay design is symmetrical and well-balanced as regards the doses employed of S.P. and T.P., the result obtained by weighting the mean response in each dose-group according to its variance will differ very little from that obtained by assuming Homoscedasticity. If the variance shows a consistent trend upwards or downwards from the lower to the upper limit of response, particularly in a series of consecutive assays, then some other transformation which will equalise the variance over the range of the assays must be sought, or alternatively the calculations must be modified so as to take into account the dependence of the variance on Y . The form of the relationship can be approximately inferred from the internal evidence of the assay (better still, of a series of assays), and, even if incorrect, the correction thus introduced will be better than no correction at all. But the bio-assayist who performs his own computations, and is usually averse from any extension of the arithmetical labour, will probably prefer to use some other transformation of his data in the first place so as to be able to make use of well-tried orthodox formulae.

B7. It is clearly necessary to ensure that the mathematical processes employed are appropriate to the assay technique, the transformations of dose and response used, and the assumptions made. Occasionally one encounters an attempt to work out the result of a 'slope-ratio' assay using the formulae applicable to 'logdose' assays! But apart from such blunders, it is not always easy to be sure that every available piece of information in the data has been utilised as fully as possible, and that the methods of statistical estimation by which the formulae are derived are of maximum efficiency in the statistical sense—that is to say, in the sense that a sample mean is a more efficient estimate of the population mean than the average of the two extreme values would be, or the estimate of the standard error derived from the squares of the deviations from the mean is more efficient than one derived from the range. It is for the statisticians to assure themselves that the computational methods they use and teach to bio-assayists are beyond reproach in these respects; there are problems here which are not yet solved. One, which strictly speaking is outside the scope of this paper, is the correct method of making allowance in *quantal* assays for the increase in precision theoretically obtained by litter-mate control, just as it is taken into account as a matter of course nowadays in quantitative assays. Yet only when the statistical techniques are as powerful as possible in the prevailing state of knowledge should the 'fiducial

limits' really be dignified with that title; figures computed by inefficient methods are approximations and unworthy of a name with such precise implications.

The utilisation of all available information implies that the data of one dose-group ought not to be omitted from the calculations for the sole reason that it makes them easier. Again, if the body-weights of the animals have been recorded at the beginning of, say, a pharmacological assay, the theoretically correct method of utilising most efficiently this information is *via* a covariance analysis [3, 5, 6, 13, 15]. The alternative practice of giving doses of drugs and the like in quantities proportional to body-weight is time-honoured but—as Bliss and Marks pointed out in 1939 [5]—illogical and inferior. There are three possibilities: the 'ideal' response, other things being equal, may be (a) independent of initial body-weight (b) linearly related to body-weight (c) related in some other way to body-weight. If (a), the administration of doses proportioned to body-weight is quite unsound. If (b), it will lead to the same results as if an analysis of covariance were performed. If (c), an analysis of covariance will give more accurate results more efficiently. Now as Mr. A. L. Bacharach has pointed out [1] the time spent in calculating and measuring out each separate dose for individual animals according to their body-weights is usually more than would be required for the extra computations in an analysis of covariance if the dose had been constant per dose-group. Moreover, in nutritional, serological, and bacteriological work, doses proportional to body-weight are almost never given. Probably the fact that the worker knows how to calculate proportional doses, but does not know how to perform an analysis of covariance, has something to do with the prejudice in favour of the former! In any event, one would expect a natural persistence of the old method (the only possibility in the days before covariance analysis had been invented) until knowledge of the newer computational techniques has become widespread and the reluctance of many bio-assayists to 'do more sums' than they can possibly help has been overcome. But it is worth pointing out that if an analysis of covariance were performed in appropriate circumstances, not only would the result of the assay be evaluated more efficiently, but some information could be obtained about the nature of the relation between responses and body-weight instead of this being assumed linear without evidence as when the 'proportioned-dose' technique is used. The same technique could and should be applied to any other factors besides body-weight known to be important; randomisation of such factors is a poor substitute.

SECTION C

THE MEANING OF THE TERMS 'ESTIMATE OF POTENCY'
AND 'FIDUCIAL LIMITS'

When all the assumptions discussed in Sections A and B have been validly made and the appropriate computations performed, there emerges the *grand dénouement*, the fulfilment of all the experimental and arithmetical labour—the result of the assay. This will normally consist of (a) the *estimate of potency*, for which we shall use the symbol R ; (b) the *fiducial limits* of R , which may be symbolised as ρ_L and ρ_U for the lower and upper limits respectively. While every bio-assayist has a more or less clear idea of what is meant by these terms, he might well have difficulty in framing their definitions in reasonably exact words, and there are many aspects of them which deserve more consideration than they are normally given, particularly the term 'fiducial limits'. Reference should be made here to an important paper by Yates [24] on this subject, from which some of the points made below are taken. It is couched, however, in phraseology rather too mathematical for the reader who is not himself a mathematician.

PART 1. THE ESTIMATE OF POTENCY

In one sense this is very easy to define, for the principal object of the assay is obviously to estimate the potency ρ as defined in Section A, and R is thus the estimate from the assay data of the 'ideal' or 'true' potency ρ . But by using different methods of calculation, different estimates of ρ could be obtained, all more or less plausible. The statistician, however, will say that there will be one estimate which is the 'best' in the sense that it is fully efficient; that other inefficient estimates are inferior; and that he can state the general rules for calculating unambiguously and with certainty this 'best' estimate R on the basis of the assumptions made.*

What he means by the 'best' estimate can be defined without much difficulty or subtlety. Let it be supposed that a long series of assays are carried out using the same S.P. and T.P., the same test subjects (or as nearly comparable test subjects as can be obtained), the same environment and the same methodology; that all the assumptions of the previous sections are exactly true for all these assays; and that to the data of each assay are applied alternative methods of calculating

*It may be in certain assays that even allowing the truth of all the assumptions listed in Sections A and B, more than one fully efficient estimate of potency can be obtained by following different rules of estimation. These will then be equally 'good'; the remainder of the present discussion is still applicable to these estimates.

an estimate of ρ . Corresponding to each method of calculation there will be obtained a long series of estimates. Then that computational method is 'best' for which (a) any one of the estimates obtained by using it would approach indefinitely closely to ρ if the assay size were increased to include more and more test subjects without limit; (b) the series of estimates has a standard deviation smaller than that of any similar series produced from the same experimental data by some other computational process. The first requirement ensures that the estimate is 'consistent' in the statistical sense; the second ensures that it is 'efficient', i.e., that the result of an assay does not deviate from the truth by more than is inherent in the experimental data, as distinct from the method used of extracting therefrom the information sought.

In practice, of course, the assumptions of Sections A and B will not be *precisely* true for any one assay, much less the entire series. Nevertheless, provided that there is no systematic bias involved in any one assumption, so that the departures from the assumptions over the series are random, that method of calculating the estimate of potency which is 'best' in the ideal case will also be best in the practical case.

The manner in which to decide the computational process which should be applied to any given set of data is outside the scope of this communication except in so far as it is discussed in Section B, Part 2 and in Section D. In the main, this is a question of general statistical theory.

PART 2. THE MEANING OF 'FIDUCIAL LIMITS'

Most bio-assayists appreciate that while the estimate of potency is the best *single* figure to quote as the result of the assay, it is not possible to assign any definite value to the probability that it is correct. In order to state the 'reliability' of the assay result quantitatively, it is necessary to give the fiducial limits corresponding to some arbitrary level of probability. (This is usually 95%, and in what follows this will be taken as the probability level for discussion, though there may often be good reasons for using some other figure.) If pressed for a definition of what his fiducial limits represent, the bio-assayist will often say that they are the limits within which it is probable, with odds of 19 to 1 on, that the true potency lies. If rather more knowledgeable than this, he will know that orthodox statistical philosophy denies the possibility of assigning a true frequency or probability distribution to an unknown population parameter.* He will then cast his definition in some such mould as this—"The fiducial limits ρ_L and ρ_U are quantities (the largest and smallest possible respectively) such that if the true

*Except in certain rather artificial problems of no practical importance.

potency were below ρ_L , then not more than $2\frac{1}{2}\%$ of all samples (i.e., assays) drawn from such a population would yield an estimate of potency as high as R , while if the true potency were above ρ_U not more than $2\frac{1}{2}\%$ of the samples would yield an estimate as low as R' . The difference between the previous inaccurate but simple definition and the more accurate but complicated definition should be noted; the latter avoids the pitfalls of inverse probability by a statement of the form 'if . . . (hypothesis about the population) then . . . (deducible consequence about the sample)' i.e., the inference is from the population to the sample. But even so, there are complexities behind this seemingly satisfactory definition which deserve examination. The logical chain of reasoning may be set out in three steps, as in the ensuing paragraphs, where the argument is simplified without any loss of generality by supposing that the samples drawn consist of n animals all in one dose-group, and that from the mean m and standard error s of the response,* statements are to be made about the 'true' response μ and standard deviation σ of the population. Extension to assays and statements about 'true' potency follows without any alteration in the reasoning.

(a) If the mean and standard deviation of the (supposedly Normal) population were both known, the probability P could be stated that the mean m of a sample of size n would lie above a given value m_U or below a given value m_L . Conversely, if it were P that was given, the corresponding values of m_U and m_L could be stated. These statements would be based on the distribution of the quantity $[(m - \mu)/\sigma] \sqrt{n}$, which is known precisely for such a population, for it is the ratio of the difference between the population and sample means to its standard error.

If a long series of samples is taken from the same population, the values of m and of the sample standard error s will in general vary from sample to sample. There will also in general be a difference between the predicted value of P as above and the actual proportion of samples having means outside the limits m_L to m_U . This difference will be due entirely to the sampling error of m ; it will have nothing to do with the sampling error of s , which does not enter into the formula by which P is calculated; and it will clearly tend to zero as the number of samples increases indefinitely, i.e., in an infinite series of samples the proportion lying outside the stated limits will be exactly P . These three statements are important in the light of what follows.

(b) When μ is known but the standard deviation is *not* known,

*The standard error s must be calculated by an efficient method, i.e. from $s^2 = \frac{\sum (x - m)^2}{(n - 1)}$.

statements of the same kind can still be made, based on the known distribution of Student's $t = [(m - \mu)/s] \sqrt{n}$. This does not involve any inferences about the probable value of σ for a given s (s of course represents the standard error calculated from the sample) and the statements made may thus appear at first to have the same logical foundations as before.

There is, however, a difference. A statement, based on the value of s for the first sample, such as 'the probability is P that the mean of a sample of size n and standard error s will lie below m_L or above m_U ' must contain the italicized words if only by implication, and will be incorrect if it does not.* But if a long series of samples is taken, s will not remain constant throughout the series. The difference between P and the actual proportion found in the long series will depend partly on the sampling error of m as before, but also on the departure of the value of s for the first sample from the mean value \bar{s} for the whole series; it will not in general tend to zero as the number of samples increases. This point does not appear to be made in any of the standard text-books; it is made by Yates [23].

(c) The further step from a known to an unknown μ does not introduce any additional logical difficulty. One can still make statements of the same kind as before provided they are prefaced by the introductory words 'If the mean of the population from which this sample was drawn be assumed to be μ , then . . .'. To reduce the number of variables, let P be fixed from now on at 0.05. Then against various hypothetical values of μ the corresponding values of m_L and m_U could be tabulated for a given sample of size n and standard error s . Once the table had been constructed, one could readily select from it two values of μ such that if the lower μ_L were correct, 97.5% of samples of size n and standard error s (note these words) would have a mean of m or less, and if the upper μ_U were correct, 97.5% would have a mean of m or more. Then, using the conventional criterion of $P \leq 0.05$, hypothetical values of μ outside the range μ_L to μ_U are regarded as disproved, and the term 'fiducial limits' is attached to μ_L and μ_U .

The point made in paragraph (b) above, that s will not in fact remain the same from sample to sample, still applies in exactly the same way. For example, if a long series of samples were drawn from a population of which the mean was μ_L and the std. deviation unknown but estimated by s (the std. error of the *first* sample), the proportion of means above m would not in general be 2.5%, nor would it even tend to that figure as the size of the series increased.

*Unless s for the first sample happens by chance to be exactly the same as the mean value \bar{s} for the whole series of samples. This colossal 'fluke' may be ignored.

We are in practice concerned with the problem, not of making one isolated statement about the fiducial limits of a population mean based on one sample, but of making a long series of such statements each based on one sample from a different population each time (e.g., successive routine assays of different Test Preparations). Suppose that in spite of the logical flaw under discussion, we always calculate fiducial limits as above and report on each experiment in the words "The mean result of this experiment is m and—for a probability of 0.95—the fiducial limits of the population mean are μ_L and μ_U ". Then these limits will sometimes be too wide (when s for the experiment concerned is above the mean value we should have obtained for s if we had performed a very large number of identical experiments) and sometimes too narrow (when the reverse is true). We shall err as often in the one direction as the other, so there will not be any systematic bias in our statements. It is *not* true, however, that 95% of them will be correct; in general, no single statement in this form will be exactly correct, except by chance, just as no single value of s obtained in an experiment is an exact estimate of the mean \bar{s} for the hypothetical series of experiments.

There is another way, however, of putting the same point which leads to more practical conclusions. Instead of saying that the limits μ_L and μ_U are sometimes further apart and sometimes closer together than the true limits corresponding to a probability of 0.95, one could say that the true probability corresponding to the limits quoted is sometimes more and sometimes less than 0.95. If a long series of similar assays were reported on in the same terms, the average probability would converge to 0.95 as the series became longer. If, therefore, the wording of the reports were to be amended to "the mean result of this experiment is m and the true value of the quantity estimated by m lies between μ_L and μ_U " it would then be true to say that the percentage of correct reports would be close to 95%.

It is curious and at first sight paradoxical that by thinking about the proportion of true assertions in a long series, we have arrived at a form of wording which appears to be as heretical an example of 'inverse probability' as the first definition of fiducial limits, which was deservedly rejected. But the reader who finds himself confused at this point should consider further the difference between the statements 'On the evidence of this isolated assay, I assert that there is a probability of 95% that the true mean lies between X and Y ' and 'On the evidence of this isolated assay, I assert that the true mean lies between X and Y , and in saying so I also assert that over a long series of similar assertions made on similar assays, I shall be correct 95% of the time.'

Definitions of the terms 'mean result' and 'fiducial limits' may now be advanced which summarise these arguments and conclusions. It is hoped that they may be helpful both to the bio-assayist with sufficient mathematics to appreciate an accurate form of wording and to those who take decisions on his reports and are interested only in the practical application of them.

The 'mean result' R is the best single estimate that can be formed, provided that certain basic assumptions are valid, of the true value of the quantity which it was the object of the assay to estimate. Because of the inherent and unavoidable errors of biological assays, the estimate R is subject to uncertainty to an extent which is measured by the 'fiducial limits' ρ_L and ρ_U .

The meaning to be attached to these figures is that if the true value of the quantity estimated were as low as ρ_L , it is very probable that further assays similar to the present one would give a mean result lower than R , and if the true value were as high as ρ_U , it is equally probable that further assays would give a result higher than R . The degree of probability referred to in these statements is not far removed from 0.95; if a long series of similar assays were performed and the fiducial limits of each calculated in the same way, the *average* probability involved in statements of the same kind would be very close to 0.95, and the longer the series, the closer the approximation would be.

For practical purposes, the quantities ρ_L and ρ_U may be regarded as limits within which it is very probable that the true value lies. If in a long series of assays the assertion is made each time that the true value lies between the fiducial limits, then 95% of these statements will be correct.

SECTION D

THE RELATIONS BETWEEN BIO-ASSAY, STATISTICS, AND REALITY

This discussion would not be complete without some further consideration of the criteria for selecting those functions of the doses and responses as actually measured which are to be used as 'metameters' in the calculations, and of the effect on the results if the assumptions enumerated in Section B are not rigidly true for the transformations chosen. It has been pointed out in Part 2 of that Section that the most to be expected of metameters in practice is that over a series of assays only random and non-significant departures from assumptions B2 to B6 are shown. Even then it is certain that there will be variants of these metameters which would have been found equally satisfactory and would have given somewhat different fiducial limits [14]. It is then a difficult problem to decide which particular transformations

should be used. It would clearly be very desirable, if it were possible, to ensure that the choice was not in any way connected with the pre-delections of the individual computer; for otherwise the calculations would lose that objectivity which is always claimed as one of the chief advantages to be gained from the use of statistics, and credence would be given to the gibe often heard from those suspicious of statistical 'manipulations', that six different statisticians, set to work on the same data, will produce six different results. This is regarded as evidence that the statistician, like a dishonest accountant, 'cooks the books' so as to produce whatever answer is most expedient. In so far as there is any substance in such a charge, it is largely because the criteria to be applied in selecting the metameters do not lead to a unique choice. Wishful thinking conjures up the impractical idea of a set of working rules which would lay down an order of preference for transformations, so that, other things being equal, $\log u$ would be tried before $1/u$; $1/u$ before $1/\sqrt{u}$; $1/\sqrt{u}$ before $1/\sqrt{u} + 5$; and so on. But it is important to realise that the selection of the transformation to be used is, as was stated earlier, a purely empirical process; there is hardly ever any theoretical reason for preferring any one to any other.

Indeed, it is possible to maintain that such questions as 'is the distribution of the response metameter precisely Normal?' have no real meaning at all; remembering that the question relates to an imaginary population of responses that might have been obtained if the experiment had been replicated *ad infinitum*, and that the only evidence is a ridiculously small sample of perhaps 20 responses, one could retort that the answer to the question is not only unknown but for ever unknowable. From this it is a small step to the suggestion that the question should never have been posed and indicates a completely unrealistic approach to the problem.

This is in fact the gravamen of a second criticism sometimes heard of statistical calculations—that before they can be made at all, so many simplifying assumptions are necessary that the operations which follow are conducted upon theoretical abstractions (e.g., a perfectly Normal variate exactly linearly related to another completely error-free variate, etc.) which bear no relation to the realities of which they are the idealised simulacra.

The sufficient answer to such criticism is that if it is illegitimate to apply to the things of this world calculations based on the entities of a theoretical world with slightly different properties, then none of the computations of physicists, engineers, and astronomers—to name no others—have any meaning at all. The computers who predict the performance of a new locomotive before it has left the drawing-board

are making assertions about a figment of the imagination, a Ghost Train, to which the real engine when built will approach more or less nearly. In the early days of such computations, lack of knowledge resulted in poor approximation, because the simplifying assumptions were *too* simple. As knowledge progresses, it becomes possible to make assumptions which although more complicated are more nearly related to the truth. The answers thus obtained are better approximations; but the calculations are much more tedious and time-consuming. Ultimately there comes a point where further approximation even if possible is not made use of; the increased precision thereby obtained is unnecessary for practical purposes and the extra labour spent on the computations would be mere waste of time. Whatever Absolute Truth may be—and the answer is metaphysical—we can all recognize the very real existence of what may be called Engineer's Truth; when the stage just described in the approximative process has been reached, the Truth has to the engineer been attained.

The processes of statistics follow an analogous path. In the present discussions several examples have been given of the possibility of making allowance in the calculations for departure from the strict truth of this or that assumption—a trend in curvature, a uniform increase in the variance of y as x increases, and so on. The added complications are quite practicable but they are also tedious. Does it really matter that the fiducial limits of an assay are evaluated as 13.9 to 21.6, though a further two hours of arithmetic would have shown that a better approximation was 14.2 to 21.4? Under what circumstances could it happen that the action taken on the result of the assay (and there is little point in an assay on which no action whatever is to be taken) would be altered thereby? And would it be realistic to suggest that because both sets of figures are only approximations based on a set of theoretical abstractions, therefore no fiducial limits should be computed at all and action should be based on the subjective opinions of some responsible (or irresponsible) person? Looked at from this point of view, the criticism now being rebutted is mere defeatism and tantamount to a denial that the pure sciences can be applied to human affairs and terrestrial phenomena.

But it is perhaps necessary to enter a *caveat* with which this philosophical—perhaps even metaphysical—dissertation may be closed. The approximations must and should be made, certainly, but the errors introduced must be random and not systematic, and above all they must not introduce bias of a kind arising from the mental characteristics and prejudices of the statistician himself. Approximations or not, the answers evaluated must be objective, and over a long series of such

answers it must not be possible to point to any constant drift away from the truth in some particular direction. Provided the deviations are random, and the fiducial limits quoted are as often too wide as too narrow, the statistician may well reply to his critics that if the action taken in all circumstances is consistently based on his reports, in the long run a smaller proportion of errors will be committed than if it were based on any other kind of criterion. He may then fairly ask them whether any other class of persons making reports on which action is taken can claim as much!

SUMMARY

Bio-assay results are usually summarised in two statements; an estimate of potency and the fiducial limits of this estimate. The truth of these statements depends on a number of assumptions which always have to be made but are never all stated explicitly when the results of the computations are presented. These assumptions fall into two groups:

Section A. Three assumptions must be made if a body of data is to be regarded as a fundamentally valid assay.

- A1. The hypothesis of validity of the experimental design.
- A2. The hypothesis of existence of a single-valued dose-response relationship.
- A3. The hypothesis of similarity of the test and standard preparations.

Section B. The statistical validity of the computation of the results depends on seven assumptions.

- B1. The assumption of relative precision in the measurement of dose.
- B2. The assumption of existence of satisfactory dose and response metameters.
- B3. The assumption of computability of the metameters.
- B4. The assumption of linearity between the dose and response metameters.
- B5. The assumption of normality of the response metameter distribution.
- B6. The assumption of homoscedasticity of the response metameter.
- B7. The assumption of efficiency in the computational methods.

Both these sets of assumptions are defined, the extent to which they are essential or may be modified is discussed, and their effect upon the methods of conducting the assay and the calculations arising from it is considered.

Section C. The precise meaning of the terms 'estimate of potency' and 'fiducial limits' is stated and analysed.

Section D. Bio-assay depends entirely upon biological research. So long as knowledge of the biological and biochemical mechanisms underlying the assay data is incomplete, bio-assay results will always be subject to errors which cannot be included in the statements of fiducial limits.

Certain criticisms affecting the legitimacy of statistical calculations as applied to bio-assay are refuted. Such calculations represent the only common-sense way of dealing with quantitative biological results. Before action is taken on the result of a bio-assay, it should always be considered how far the assumptions enumerated above are sufficiently well-founded in that particular assay to be acceptable for practical purposes.

REFERENCES

1. Bacharach, A. L. Unpublished correspondence with E. C. Wood.
2. Bliss, C. I. *Journal of the American Pharmaceutical Association*, 29: 465, 1940.
3. Bliss, C. I. *Journal of the American Statistical Association*, 35: 498, 1940.
4. Bliss, C. I. *Industrial and Engineering Chemistry*, 13: 84, 1941.
5. Bliss, C. I. and Marks, H. P. *Quarterly Journal of Pharmacy and Pharmacology*, 12: 82 and 182, 1939.
6. Burn, J. H. *Physiological Reviews*, 10: 146, 1930.
7. Cochran, W. G. *Biometrics*, 3: 22, 1947.
8. Emmens, C. W. *Principles of Biological Assay*, London: Chapman and Hall, 1948.
9. Fieller, E. C. *Journal of the Royal Statistical Society, Suppl.*, 7: 1 and 50, 1940.
10. Fieller, E. C. *Analyst*, 72: 37, 1947.
11. Finney, D. J. *Nature*, 153: 284, 1944.
12. Finney, D. J. *Journal of the Royal Statistical Society, Suppl.*, 9: 46, 1947.
13. Finney, D. J. *Journal of Hygiene*, 45: 397, 1947.
14. Finney, D. J. *Biometrics*, 1949, —.
15. Fisher, R. A. *The Design of Experiments*, London: Oliver & Boyd, 4th edition, 1947.
16. Gaddum, J. H. *Biochemical Journal*, 25: 1113, 1931.
17. Gaddum, J. H. *Medical Research Council Special Report Series*, no. 183, 1933.
18. Geary, R. C. *Biometrika*, 34: 209, 1947.
19. Irwin, J. O. *Journal of the Royal Statistical Society, Suppl.*, 4: 1, 1937.
20. Rasch, G. *Biometrics*, 3: 173, 1947.
21. Thompson, W. H. *Biometrics*, 4: 97, 1948.
22. Trevan, J. W. *Proceedings of the Royal Society, B101*, 483, 1927.
23. Wood, E. C. *Analyst*, 71: 1, 1946.
24. Yates, F. *Proceedings of the Cambridge Philosophical Society*, 35: 579, 1939.

A BIOLOGICAL ASSAY OF TUBERCULINS

R. A. FISHER

University of Cambridge

A FEW YEARS AGO it was decided to carry out under the authority of the Agricultural Research Council a specific test, using bovine subjects, of the relative potency of two tuberculin preparations, which may be designated as Standard and Weybridge respectively. Such a test constitutes essentially a biological assay of the tuberculins, and a report of its results may be of interest, since little seems to be known of the statistical problems involved in the use of the tuberculin reaction for such a purpose.

For the test ten herds in different parts of England were used, and from each, twelve cows were chosen and assigned to four treatment groups, each group thus receiving three cows of each of ten herds. The groups differed only in the sites at which the tuberculin was applied. The treatments applied at each site were

<i>A</i>	Standard	0.1 mgm.
<i>B</i>	Standard	0.05 mgm.
<i>C</i>	Weybridge	0.05 mgm.
<i>D</i>	Weybridge	0.025 mgm.

The sites of application, four on each side of the neck, were numbered from one to eight in such a way that numbers five to eight on the left side corresponded with numbers one to four on the right. At each site, the measurement made was a thickening of the skin observable in a set number of hours after intradermal injection of the tuberculin. The treatments of the four classes of cow are set out in the following table:

Sites	Treatment Class			
	1	3	2	4
3 and 6	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
4 and 5	<i>B</i>	<i>A</i>	<i>D</i>	<i>C</i>
1 and 8	<i>C</i>	<i>D</i>	<i>A</i>	<i>B</i>
2 and 7	<i>D</i>	<i>C</i>	<i>B</i>	<i>A</i>

The following sections will incorporate parts of the report made by the author to the Agricultural Research Council in March 1944. The interest of the matter to biometricians lies in the fact that tuberculin readings present in an acute form the need to develop *ad hoc* an appropriate theory of errors. This need is often present and sometimes unrecognised in other types of biological response. The preliminary investigations, by which a theory of errors appropriate to these readings was built up, may therefore be of assistance to workers with other material.

It will be seen that the method judged by these tests, and verified *a posteriori* to be appropriate to the material, is essentially that of χ^2 analysis as ordinarily used with observations of frequency. This was adopted not only because it works well, but because it lies ready made to the hand of the statistician. I do not think it is the only mode of analysis which could have been usefully applied. Indeed the Eulerian distribution

$$\frac{1}{p!} x^p e^{-x} dx$$

having, for variable p , variances proportional to the means and giving exhaustive simultaneous estimation based only on the arithmetic and geometric means of each sample, would seem to supply an equally effective mode of approach and one which it would be of considerable mathematical interest to develop. I cannot, however, imagine that it should give a different answer to the practical question at issue.

I should add, what was not known to me when I wrote the report, that careful comparative tests with guinea pigs, well designed and of high precision, gave in fact a ratio of 0.9 instead of 2.2 for the two materials. They must, therefore, in reality be qualitatively different, although there is no indication of this within the scope of the bovine test.

The analysis of the experiment designed to assay the potency of Weybridge P.P.D. H₇ Tuberculin encountered two difficulties:

(a) That arising from the very great variation in the reaction of different cows. This of course had been foreseen as inevitable in unselected material, and it had been proposed that the series of trials first carried out should be regarded in one aspect as a means of selecting animals of uniformly high reactivity, a panel of which could be used for a more accurate assay.

As this had been found impracticable, it was necessary to utilize data involving the full variation in reactivity of unselected material.

(b) It was anticipated that equivalent reactions would be obtained from like sites on opposite sides of the same neck. The data available from the repetition of the test on 120 cows in all show that no such similarity is to be relied on, but that significant differences between Right and Left occur, and that these differences are strikingly different at the four chosen sites. In fact the data have to be examined as if each of the eight test-points on each animal had its own characteristic sensitivity.

In consequence of these two drawbacks the methods of reduction which we had hoped to use appeared on examination to be quite inadequate for the purpose of combining the information available from the different parallel sets of animals. In forming a judgement as to the manner in which such factors as tuberculin-potency interact with the differences in sensitivity of different animals and of different sites, the most valuable information is provided by the fact that on each animal certain pairs of sites, namely 1 and 8, 2 and 7, 3 and 6, 4 and 5, are invariably treated alike, although the actual treatments used on these pairs of sites are varied for animals of the four different classes.

Preliminary analysis seemed to indicate that the difference in response at two sites on different cows was proportional to the sensitivity of the cows, and that the difference in response to different treatments was proportional to the general average response which such treatments provoke. This, so far as it may prove to be true, is a most valuable generalisation. Together with a second observation, namely that the variance to be ascribed to any observation, whether owing to the individuality of the animal or to errors of measurement, is approximately proportional to the magnitude of the measurement to be expected, it does allow of a rational and comprehensive form of analysis.

To demonstrate the approximate truth of these views, the animals in each class were divided, using the total reaction at 48 hours, into 4 groups of reaction-intensity: thus, of the 30 animals in class 1, four, giving reaction at 48 hours of 0-19mm. in all at the 8 sites, are of the lowest class of reactors (α), nine give reactions of 20-49mm., i.e. on the average $2\frac{1}{2}$ to 6mm., (class β), eleven give total reactions of 50-79mm. i.e. 6-10mm. on the average, (γ), and six give total reactions of 80 or more, (δ).

Taking, for example, sites 1 and 8, which with these 30 cows both receive treatment *C* (Weybridge 0.05mgm), if a and b are the measurements observed at any stage, e.g. 72 hours, we can calculate $[(a - b)^2]/(a + b)$ for each cow, and for any group of cows $[(A - B)^2]/(A + B)$ where A and B are the sums of a and b . Then for variation in the ratio

of measurement at site 1 to that at site 8 among the 4 cows of the lowest sensitivity-class, one has three degrees of freedom, yielding

$$S\left\{\frac{(a-b)^2}{a+b}\right\} = 6.7777$$

$$\frac{(A-B)^2}{A+B} = \underline{2.8824}$$

leaving 3.8953

as the contribution of these three degrees of freedom. Since there are three other pairs of sites equally comparable on each cow, also with nearly equal total reaction, one can in this way make up 12 degrees of freedom, obtaining the total of 6.9948, measuring variation of the same sort within homogeneous material. The three other classes of cow, in which these same sites receive treatments, *A*, *D*, and *B* respectively, bring up the total degrees of freedom to 108, with a total sum of squares of 40.8580, and a mean-square measured in this way for the least responsive class of cows (α) of 0.38mm.

The point of this procedure is the comparison it allows between cows of very different absolute sensitivity. For the four classes of cows chosen one has the results shown below:

Reactivity-class		Degrees of Freedom	Sum of Squares	Mean-Square
α ,	31 cows	108	40 8580	0 3783
β ,	43 cows	156	70 0597	0 4491
γ ,	35 cows	124	50 8053	0 4097
δ ,	11 cows	28	14 3365	0.5120
120 cows		416	176 0595	0 4232

Measured in this way, therefore, the gross heterogeneity between cows of different sensitivity-classes has practically disappeared, and the contributions of unequal numbers of cows in these classes to the evidence may be satisfactorily weighted. Further it appears that the ratio of reaction-measurement at two comparable sites is nearly the same whatever treatment these sites receive. For each sensitivity-class of cow, twelve degrees of freedom have been excluded from the analysis above, representing possible differences of this kind. For the four sensitivity-classes, these are:

Class	Degrees of Freedom	Sum of Squares	Mean-Square
α	12	6.5598	0.5466 mm.
β	12	5.9111	0.4522
γ	12	1.7141	0.3862
δ	12	12.8481	1.0290
	48	26.5331	0.5528 mm.

Apart from the slight suggestion that in the most sensitive class of cows some heterogeneity in the site-ratio has been introduced by varying the tuberculin used, these figures show that there is little danger of being misled if the data are treated as though the ratio of the response at different sites, both in different cows and to different tuberculins, were a constant property of those sites. This is important, since of the four pairs of sites treated alike, three (namely 1 and 8, 3 and 6, 4 and 5) all show significantly unequal response in the aggregate examined. Finally not only is the variation homogeneous within groups of cows showing very varying sensitivity to tuberculin, but the ratio of response in the four classes of cows chosen for their different sensitivity is also the same. For this we have three degrees of freedom for each pair of sites, or twelve in all:

VARIATION AMONG DIFFERENT SENSITIVITY-CLASSES $\alpha, \beta, \gamma, \delta$

Degrees of Freedom	Sum of Squares	Mean-Square
12	5.0866	0.4239

On the basis of this preliminary investigation, which has been set out in detail above for readings at 72 hours, the problem of estimating the proportionate increase in swelling measured produced (a) by doubling the quantity of tuberculin, and (b) by replacing a given amount of Standard tuberculin by half the quantity of Weybridge 10, becomes tolerably straightforward.

The method used in the original report, although substantially accurate in the results it gave, was not well suited as a methodological model, and may be replaced for our present purposes by one of equivalent accuracy and perhaps greater clarity.

Taking, for example, the data for readings at 48 hours, and adding

together readings at the two sites treated alike and on the 30 cows treated alike, the aggregate results of the test may be expressed by the following 4×4 table, to which is appended on the right a key to the treatments used in the form of a non-cyclic Latin Square.

Sites	Cow Class								
	I	III	II	IV					
3 + 6	454	249	349	249	1301	A	B	C	D
4 + 5	408	322	312	347	1389	B	A	D	C
1 + 8	523	268	411	285	1487	C	D	A	B
2 + 7	364	283	266	290	1203	D	C	B	A
	1749	1122	1338	1171	5380				

Treating these aggregate measurements as quasi-frequencies, the data now have a form closely similar to that which arises with a three-point linkage test in genetics, in which also we have 16 observable frequencies, classifiable in three orthogonal categories, assigned arbitrarily to the rows, columns and letters. In such a case, for example, we have typically four different triple heterozygotes used as parents and assigned to the four rows, four modes of gamete formation (crossover classes) assigned to the four columns, and four pairs of complementary genotypes distinguishable associated with the four letters of the square.

If, as sometimes happens, these pairs of complementary genotypes are not equal in viability, the frequencies to be expected in the sixteen entries will be affected not only by factors representing modes of gamete formation and abundance of material from the four possible sources, but by a third unknown set of factors representing relative viabilities.

The statistical problem will then consist in assigning sixteen expectations to the sixteen cells of the table, each expectation being the product of three appropriate factors, all of them unknown.

An examination of this statistical problem shows that the solution of maximum likelihood is such that the sums, by rows, by columns, and by letters, of the expectations are equal to the corresponding sums of the observed frequencies. This is a statistical solution of the utmost simplicity, although the algebraic problem of constructing expectations fulfilling these marginal conditions, and the condition of being triple products, seems to be one of some intricacy. I have elsewhere discussed certain approximate methods of approach.¹

The tuberculin data are in one respect slightly simpler than the

¹R. A. Fisher (1949) Note on the test of significance for differential viability in frequency data from a complete three point test *Heredity* 3, 2, 215-219.

corresponding genetical problem, for in this case it is to be presumed, unless the data indicate otherwise, that the effect of doubling the dose is the same whichever of the two tuberculins is used, i.e. that the factors corresponding with the letters A , B , C and D shall be in proportion. This circumstance opens the way to an effective approximate estimate of these factors.

It will be noticed in the symbolic square that the four quarters are constituted by 2×2 Latin Squares such as

$$\begin{array}{cc} A & B \\ B & A \end{array}$$

so that the ratio $A : B$, representing the ratio of the readings for double and single injections, can be consistently estimated from the product ratio of the four observed total measurements, i.e. from

$$\sqrt{454 \times 322/408 \times 249}$$

In practice it is most convenient to work with natural logarithms, so that we have

Treatment	Total Measurement	Natural Logarithm
A	454	1.51293
B	408	-1.40610
A	322	1.16938
B	249	-.91228
	$A : B$.36393
		.18196

The weight of this logarithmic estimate is ("The Design of Experiments", Section 70) the harmonic mean of the four frequencies, namely 339.69.

Taking in turn the three other similar included 2×2 squares, and remembering that the ratio $C : D$ is to be presumed equal to that of $A : B$, we have, in the four cases

	Log Ratio	Weight
$A : B$.18196	339.69
$A : B$.22624	304.19
$C : D$.20844	335.45
$C : D$.22197	308.44

from which we have the estimate of the weighted mean .20890, for the effect expressed as the natural logarithm of the measurement of doubling the tuberculin dosage.

It may also be seen that four more 2×2 Latin Squares, in this case overlapping, are available to estimate the ratio $A : C$ or $B : D$ for which, using again natural logarithms, we have the estimates

	Log Ratio	Weight
$A : C$.01102	424.94
$A : C$	-.02517	308.42
$B : D$	-.02269	328.87
$B : D$.03075	261.91

the weighted mean being in this case $-.00188$. It will be noticed at this stage that the estimates are showing a remarkable consistency.

The two estimations carried out above answer the practical question of the enquiry by assigning relative performance to the single and double doses of the two tuberculins used and show, in fact, that the Weybridge material was effectively a little more than twice as potent as the Standard. Questions of precision can, however, only be answered by constructing the expectations corresponding to the measurements observed. An approximate method of doing this, appropriate to cases like the present, in which all cells of the square are well occupied, is shown below.

We have the measurements of logarithmic relative potency

B	Standard single	0.0000
A	Standard double	0.2089
D	Weybridge half	0.0019
C	Weybridge single	0.2108

The antilogarithms of these give factors appropriate to the four treatments. Dividing the observed frequencies by these factors we have the adjusted frequencies

368.412	249.	282.673	248.532	1148.617
408.	261.297	311.413	281.053	1261.763
423.604	267.496	333.518	285.	1309.618
363.316	229.216	266.	235.329	1093.861
1563.332	1007.009	1193.604	1049.914	4813.859

From the margins we can reconstruct the table so that the rows and columns are in strict proportion

373.0208	240.2787	284.8014	250.5161
409.7657	263.9476	312.8561	275.1936
425.3070	273.9584	324.7219	285.6307
355.2385	228.8243	271.2246	238.5736

Each value may now be multiplied by the appropriate treatment factor, so as to give an approximate expectation.

459.6817	240.2787	351.6274	250.9876	1302.5754
409.7657	325.2684	312.4430	339.7652	1387.2423
525.113	274.4740	400.1619	285.6307	1485.3679
355.9071	282.5158	271.2246	293.9995	1203.6470
1750.4558	1122.5369	1335.4569	1170.3830	5378.8326

Since these do not give exactly the original total, they may be reduced to the correct total, as in the following table.

459.781	240.331	351.704	251.042	1302.858	-1.858
409.855	325.339	312.511	339.839	1387.544	+1.456
525.215	274.534	400.249	285.693	1485.691	+1.309
355.984	282.577	271.283	294.063	1203.907	-0.907
1750.835	1122.781	1335.747	1170.637	5380.000	
1749	1122	1338	1171		
-1.835	-0.781	+2.253	+0.363		

The marginal totals of this table of expectations, although not exactly equal to those of the observations on which they are based, are good approximations to these. Thus the column totals each of about 1300mm. have discrepancies -1.8, -0.7, +2.3, +0.4mm. only. With the rows the largest discrepancy is only -1.9mm., and with the letters (treatments) we have

	Expected	Observed	
A	1479.432	1477	-2.432
B	1207.162	1208	+0.838
C	1499.335	1503	+3.665
D	1194.071	1193	-1.071

Thus our method, though only tentative and approximate, can be seen after the event to have given a very satisfactory approximation to the ideal fitting required. Owing to the importance of this type of problem in genetics, and the probability of further analogous cases in biological assay, the problem of making a sufficiently rapid and sufficiently accurate fitting of this kind seems to deserve further study.

Given sufficiently good expectations, we can calculate the ingredients $(a - m)^2/m$, the sum of which supplies the analogue to χ^2 for the residual seven degrees of freedom.

$(\alpha - m)^2/m$				
.0093	.1555	.2888	.0017	1.4126 mm.
.1805	.0006	.1029	.0561	
.0727	.3127	.0208	.0166	
.0084	.0343	.0008	.1509	
.2709	.5031	.4133	.2253	
d.f.				
7	1.4126	.2018 mm.		
12	2.92564	.2438 mm.		

In millimetres this comes to 1.4126, with a mean square .2018mm. only. This value is in good agreement with that obtained by contrasts between the aggregate readings on sites treated alike, which for twelve degrees of freedom gives 2.92564, or .2438mm. as the average value.

These values are rather surprisingly less than those obtained directly in the preliminary test set out above. The indications of precision available from individual readings were, therefore, recalculated more exactly, treating each set of three cows of the same herd and treatment as a 3×8 frequency table, giving 21 degrees of freedom within the herd, and each set of trios, one from each of ten herds for a given treatment, thus supplying 63 degrees of freedom between herds. Owing to a few cows giving completely zero readings, we have not quite the full number of degrees of freedom available, but using the same readings, i.e. at 48 hours, as those used in the illustration above, we have

	Degrees of Freedom	Sum of Squares	Mean Square
Within herds	539	242 7791 mm.	.4504 mm.
Between herds	252	144.1697 mm.	.5712 mm.

It is a puzzling feature, and one that I do not understand, that the comparisons used in the final aggregates should agree so appreciably more closely than do the individual readings on which they are based.

The ratio of potency of equal weights of the two tuberculins were estimated, for the readings at the three periods used, to be as follows:

Period	48 hours	72 hours	96 hours
Estimated ratio	2.009	2.141	2.172

TABLE 1
PERCENTAGE INCREASE IN MEASUREMENT (MEASURED LOGARITHMICALLY)
ESTIMATED INDEPENDENTLY FOR EACH HERD

	48 hrs.	72 hrs.	96 hrs.
DOUBLE v. SINGLE DOSE			
Kent	30.7	33.8	26.2
Cheshire A	30.0	28.2	30.4
Berkshire	25.1	32.7	25.0
Lancashire B	27.6	20.5	18.4
Cheshire B	17.8	21.8	20.7
Durham	18.4	20.0	18.3
Cambridge A	18.9	17.3	15.7
Essex	15.7	18.3	16.4
Cambridge B	17.5	15.3	- 2.0
Lancashire A	15.2	6.9	- 2.7
Weighted mean	21.1	21.5	16.4
WEYBRIDGE v. STANDARD TUBERCULIN.			
Kent	3.2	3.3	5.4
Cheshire A	- 2.1	1.7	- 0.1
Berkshire	6.8	7.5	11.2
Lancashire B	11.3	7.5	8.4
Cheshire B	-10.8	7.6	0.7
Durham	11.3	3.7	- 0.9
Cambridge A	- 4.5	6.3	7.1
Essex	- 2.4	4.1	- 9.3
Cambridge B	2.6	4.2	6.8
Lancashire A	- 3.0	5.9	-10.0
Weighted mean	0.48	3.19	3.03
<i>Relative potency of equal weight of Tuberculin.</i>			
Weybridge v. Standard,	2.032	2.217	2.274
with fiducial limits	2.341	2.505	2.727
	1.764	1.961	1.897
Estimate from aggregated data .	2.009	2.141	2.172

To examine the consistency of the differential responses on which these estimates are based, and to obtain an appropriate standard error and fiducial limits for the estimates, a parallel process was applied to the ten constituent herds individually. (The original report then discusses individual herds in detail at the different periods at which the swellings were read.) The herd values are shown in Table 1. It is upon these that the fiducial limits have been based.

Table 2 gives the relative performance at the eight sites. Of these the most forward (1, 5) are the most sensitive, and the hindermost (3,7) are least so. It is obvious that there is no consistency in the differences between Right and Left.

TABLE 2
PROPORTIONATE RESPONSE AT EACH SITE

	48 hours	72 hours	96 hours
Site			
1	1.141	1.131	1.143
2	.924	.931	.925
3	.931	.894	.892
4	.993	.990	1.017
5	1.100	1.115	1.107
6	.975	.990	.959
7	.895	.928	.930
8	1.040	1.029	1.027
	7.999	8.000	8.000

The complete data from the experiment are shown in Table 3.

SUMMARY

The above details and the result of the experiment reported have been published at the present time: partly in illustration of the fact that each type of reading which arises in biological assay deserves and may require the development for it of an appropriate theory of errors; secondly because previous work with tuberculin readings seems to have given no idea as to how they can be quantitatively interpreted; and thirdly because the precision of such readings regarded as a biological assay seems to have been much underrated.

TABLE 3
TABLE OF INDIVIDUAL RESPONSES AT 48, 72, 96 HOURS
(Data relate to 10 farms, 4 treatment classes at each, 3 cows per class).

Treatment class	Site	Fuber- culin	Kent									Cheshire A								
			Cow 1			Cow 2			Cow 3			Cow 1			Cow 2			Cow 3		
			45	72	96	48	72	96	48	72	96	48	72	96	48	72	96	48	72	96
1	1	A	16	18	11	0	4	4	21	28	21	10	12	14	10	10	10	8	8	6
	2	B	12	14	10	2	2	2	14	14	12	5	6	7	6	7	5	6	7	5
	3	C	4	6	5	7	6	4	23	24	17	6	8	9	6	8	7	6	7	4
	4	D	14	15	12	4	3	3	11	12	10	6	8	9	7	7	6	4	6	4
	5	D	6	9	9	4	2	1	14	10	12	8	10	11	9	10	8	7	7	5
	6	C	8	12	10	9	8	6	19	21	15	10	12	13	9	11	10	6	6	4
	7	B	3	5	8	2	2	2	13	16	12	7	10	10	7	9	8	5	6	3
	8	A	10	15	13	4	5	5	20	24	18	9	12	13	10	10	9	8	8	5
2	1	C	8	9	4	3	8	9	18	22	14	9	10	12	14	13	11	12	11	11
	2	D	4	5	4	2	3	7	7	9	4	5	8	7	9	8	8	8	7	7
	3	A	8	8	4	2	2	3	10	13	9	6	10	8	11	13	9	10	9	9
	4	B	5	6	3	1	2	3	9	12	8	5	8	7	12	12	10	8	8	7
	5	B	5	8	4	2	3	5	7	7	3	6	9	8	13	13	11	11	11	9
	6	A	5	6	4	3	3	5	9	12	7	8	11	9	12	12	9	10	10	9
	7	D	3	4	4	3	3	3	10	12	7	7	10	8	7	7	6	10	12	9
	8	C	7	8	6	1	1	1	9	17	11	7	10	9	12	12	10	11	16	13
3	1	B	1	1	1	3	3	5	2	4	4	5	6	5	1	1	2	0	1	1
	2	A	1	1	1	1	2	3	2	3	3	6	7	3	1	1	1	0	0	0
	3	D	1	0	1	2	3	0	2	2	2	2	4	4	0	0	0	0	0	0
	4	C	4	2	3	1	3	6	5	5	6	8	8	6	1	1	1	0	0	0
	5	C	0	1	1	4	6	6	4	5	5	5	8	7	0	0	1	0	0	1
	6	D	0	0	0	2	4	0	3	4	3	3	5	4	0	0	1	0	0	0
	7	A	1	1	1	3	4	4	3	4	4	4	4	3	1	1	2	0	0	1
	8	B	1	0	1	4	4	5	3	2	2	5	5	3	0	1	0	0	0	0
4	1	D	1	4	2	7	8	7	3	4	5	5	7	6	2	2	3	4	7	6
	2	C	2	4	3	8	11	7	3	4	4	4	4	5	1	0	2	5	8	7
	3	B	2	6	3	7	7	2	8	8	4	4	6	4	0	0	2	0	0	0
	4	A	3	6	4	8	11	6	3	4	6	6	8	7	0	0	3	5	8	8
	5	A	3	7	7	11	12	7	7	7	6	6	7	6	1	1	4	5	9	8
	6	B	3	5	2	7	10	6	3	4	4	4	5	5	0	1	3	3	5	5
	7	C	3	4	5	8	10	6	3	5	5	4	4	5	0	0	3	4	7	6
	8	D	1	3	2	7	9	6	2	4	4	3	4	4	0	0	3	2	4	4

TABLE 3—Continued
TABLE OF INDIVIDUAL RESPONSES AT 48, 72, 96 HOURS
(Data relate to 10 farms, 4 treatment classes at each, 3 cows per class).

Treatment class	Site	Tuberculin	Berkshire												Lancashire B																							
			Cow 1						Cow 2						Cow 3						Cow 1						Cow 2						Cow 3					
			48		72		96	48		72		96	48		72		96	48		72		96	48		72		96	48		72		96						
1	1	A	0	6	6	1	4	4	0	10	8	4	9	9	2	7	8	4	10	11	2	7	4	10	4	10	11	2	7	8	4	10	11					
	2	B	0	4	4	2	3	2	1	1	1	1	8	8	7	1	4	4	3	5	7	1	4	3	5	7	1	4	4	4	9	11						
	3	C	1	6	7	2	6	6	4	7	7	4	8	7	4	4	6	4	4	9	11	4	4	4	9	11	4	4	4	9	11							
	4	D	1	6	5	2	4	4	3	0	7	4	6	7	1	4	5	3	8	12	1	4	3	8	12	1	4	3	8	12								
	5	D	2	7	7	2	4	5	4	6	8	8	6	6	1	3	5	3	11	10	1	3	11	10	1	3	11	10	1	3	11							
	6	C	1	8	7	2	4	4	2	8	8	4	7	8	3	6	6	7	12	11	1	6	7	12	11	1	6	6	7	12								
	7	B	0	1	3	2	3	4	0	0	2	2	5	7	1	3	4	2	5	7	1	3	4	2	5	7	1	3	4	2	5							
	8	A	2	5	4	2	5	4	3	8	9	3	6	8	2	4	6	4	5	9	2	4	5	9	2	4	5	9	2	4	5							
2	1	C	10	13	12	17	23	23	6	8	7	4	8	11	5	7	7	10	11	15	5	7	10	11	15	5	7	10	11	15	5	7						
	2	D	8	9	9	18	18	14	2	3	5	3	4	8	6	8	9	6	10	8	6	8	10	6	10	8	6	8	10	6	10							
	3	A	9	12	12	16	18	14	3	5	5	7	9	13	7	7	7	9	13	10	7	7	9	13	10	7	7	9	13	10								
	4	B	9	9	10	17	12	20	6	6	4	3	7	11	4	6	6	4	10	10	4	6	4	10	10	4	6	4	10	10								
	5	B	12	13	12	14	16	16	3	7	5	4	7	15	2	4	3	4	12	12	2	4	3	4	12	12	2	4	3	4								
	6	A	9	9	9	19	22	20	2	5	4	12	12	13	7	9	8	8	14	12	7	9	8	14	12	7	9	8	14									
	7	D	7	7	7	9	13	18	3	3	3	4	6	10	7	8	7	6	9	9	7	8	7	6	9	9	7	8	7	6								
	8	C	14	16	15	16	17	18	6	6	4	5	7	11	6	6	6	8	11	10	6	6	8	11	10	6	6	8	11	10								
3	1	B	0	1	5	9	9	0	2	9	8	3	9	9	10	11	12	6	10	10	10	11	12	6	10	10	10	11	12	6	10							
	2	A	2	4	6	8	7	6	1	6	7	3	8	10	11	13	15	5	9	7	3	7	5	9	7	3	7	5	9	7	3							
	3	D	2	2	7	10	9	8	2	8	8	2	8	10	10	12	11	4	7	7	10	12	11	4	7	7	10	12	11	4	7							
	4	C	2	2	7	10	8	8	2	9	13	4	10	12	15	20	18	5	6	8	15	20	18	5	6	8	15	20	18	5	6							
	5	C	3	4	5	17	10	18	2	9	10	4	12	16	12	18	18	5	6	8	12	18	18	5	6	8	12	18	18	5	6							
	6	D	1	2	4	2	5	4	0	2	8	4	10	11	10	9	13	4	7	7	10	9	13	4	7	7	10	9	13	4	7							
	7	A	2	3	6	10	10	8	3	7	8	1	11	11	3	12	12	8	7	6	3	12	12	8	7	6	3	12	12	8	7							
	8	B	1	1	5	13	11	10	0	4	5	5	11	11	3	5	6	2	6	6	3	5	6	2	6	6	3	5	6	2	6							
4	1	D	6	6	4	1	4	5	1	7	6	5	12	11	2	4	5	3	7	0	2	4	5	3	7	0	2	4	5	3	7							
	2	C	6	6	4	1	4	4	3	9	6	7	12	11	3	7	9	3	9	7	3	7	9	3	9	7	3	7	9	3	9							
	3	B	5	5	4	1	2	3	4	9	7	7	8	6	1	3	4	3	5	6	1	3	4	3	5	6	1	3	4	3	5							
	4	A	7	7	4	1	2	2	4	15	13	5	10	8	2	5	6	4	8	9	2	5	6	4	8	9	2	5	6	4	8							
	5	A	8	6	4	1	6	4	4	14	8	4	12	10	2	5	6	4	7	6	2	5	6	4	7	6	2	5	6	4	7							
	6	B	5	5	4	1	1	3	3	8	4	0	11	8	2	6	5	4	4	7	6	2	6	5	4	4	7	6	2	6	5							
	7	C	6	6	4	1	1	4	3	8	5	5	9	7	2	0	7	5	6	6	2	0	7	5	6	6	2	0	7	5	6							
	8	D	9	9	4	1	1	2	1	8	7	7	14	9	2	5	6	3	8	7	2	5	6	3	8	7	2	5	6	3	8							

TABLE 8—Continued
TABLE OF INDIVIDUAL RESPONSES AT 48, 72, 96 HOURS
(Data relate to 10 farms, 4 treatment classes at each, 3 cows per class).

Treatment class	Site	Tuber-culin	Cheshire B												Durham																							
			Cow 1						Cow 2						Cow 3						Cow 1						Cow 2						Cow 3					
			48	72	96	48	72	96	48	72	96	48	72	96	48	72	96	48	72	96	48	72	96	48	72	96	48	72	96									
1	1	A	7	10	6	13	18	9	12	14	14	8	2	4	0	1	3	11	10	13																		
	2	B	5	8	5	7	9	9	7	8	10	5	2	4	1	0	2	9	9	11																		
	3	C	6	9	5	6	9	9	8	9	11	10	9	2	4	1	1	7	7	11																		
	4	D	6	9	6	7	9	9	9	10	10	10	2	4	0	1	1	10	11	11																		
	5	D	8	12	9	9	12	11	11	11	10	9	2	2	0	1	1	9	12	11																		
	6	C	6	11	8	7	10	10	8	10	9	9	2	3	0	2	2	6	8	9																		
	7	B	6	10	5	8	10	10	9	11	9	11	2	1	0	1	1	4	9	8																		
	8	A	9	13	6	12	18	15	9	11	10	10	2	4	1	2	2	11	14	13																		
2	1	C	13	14	14	14	13	10	9	9	8	2	0	1	11	10	11	1	1	2																		
	2	D	11	11	9	9	10	7	5	6	5	3	2	2	1	5	9	8	1	1	2																	
	3	A	14	14	12	11	11	7	6	8	6	2	0	2	2	2	1	0	0	0																		
	4	B	15	15	10	11	10	6	5	8	6	3	3	2	5	6	7	0	0	0																		
	5	B	22	23	17	16	16	10	7	8	6	2	2	4	9	3	5	0	1	1	1																	
	6	A	16	13	13	14	13	9	7	8	6	2	2	2	5	3	6	0	0	0																		
	7	D	9	12	9	12	11	9	5	8	6	2	1	1	6	3	4	0	0	0																		
	8	C	20	20	15	13	14	10	0	10	7	5	5	5	14	11	8	0	0	0																		
3	1	B	1	1	1	14	14	12	0	0	0	10	14	15	4	12	9	1	1	1																		
	2	A	1	2	2	9	11	8	0	0	0	11	17	14	4	8	5	0	0	0																		
	3	D	1	1	1	6	6	6	0	0	0	5	8	9	4	5	5	0	1	0																		
	4	C	0	1	1	11	13	12	0	0	0	8	15	18	7	12	11	1	1	1																		
	5	C	2	3	3	11	14	12	0	0	0	13	16	17	8	12	8	1	1	1																		
	6	D	0	0	1	7	8	7	0	0	0	9	13	13	6	8	6	1	1	1																		
	7	A	1	1	2	5	6	6	0	0	0	13	21	20	6	14	8	1	1	0																		
	8	B	1	0	0	7	9	7	0	0	0	9	15	16	0	11	9	1	0	0																		
4	1	D	2	3	3	1	0	0	3	7	9	0	2	0	8	8	0	3	4	5																		
	2	C	2	4	4	1	1	1	6	8	9	2	4	4	5	9	5	6	6	7																		
	3	B	1	3	3	0	0	0	3	6	7	3	4	4	4	6	5	3	3	5																		
	4	A	4	7	5	0	0	1	6	8	11	2	5	3	4	9	5	5	6	7																		
	5	A	4	7	5	0	0	1	5	8	10	4	6	5	4	6	7	6	8	8																		
	6	B	2	6	4	1	0	0	6	7	9	3	4	4	5	5	3	4	6	6																		
	7	C	4	7	5	0	0	0	5	10	12	5	5	6	7	8	5	5	5	8																		
	8	D	2	3	2	1	1	1	4	7	8	2	6	5	6	3	4	5	5	6																		

TABLE 3—Continued
TABLE OF INDIVIDUAL RESPONSES AT 48, 72, 96 HOURS
(Data relate to 10 farms, 4 treatment classes at each, 3 cows per class).

Treatment class	Site	Tuber- culin	Cambridge A									Essex											
			Cow 1			Cow 2			Cow 3			Cow 1			Cow 2			Cow 3					
			48	72	96	48	72	96	48	72	96	48	72	96	48	72	96	48	72	96			
1	1	A	4	7	3	11	11	7	3	2	2	4	5	3	4	12	7	9	12	7	9	12	7
	2	B	4	5	3	8	12	0	3	0	2	2	5	4	3	8	2	2	2	2	3	3	
	3	C	5	6	4	10	12	7	3	2	2	4	5	4	4	10	6	4	6	4	6	4	
	4	D	3	5	3	8	9	8	3	0	1	4	5	4	4	11	4	3	3	6	3	4	
	5	D	5	6	4	12	14	10	3	3	2	4	5	4	4	8	7	4	4	8	4	4	
	6	C	3	5	4	9	12	8	4	3	2	2	2	2	6	12	7	1	2	2	2	2	
	7	B	4	4	4	7	0	6	2	1	2	3	2	2	5	8	5	5	7	7	11	11	
	8	A	4	6	4	13	14	10	3	0	1	3	2	2	5	10	5	5	8	5	8	9	
2	1	C	4	6	5	5	6	6	24	21	10	7	12	12	10	12	6	9	10	7	5	7	
	2	D	4	6	5	1	3	3	16	13	9	5	8	8	7	8	4	6	7	6	6	6	
	3	A	5	4	4	3	4	4	22	18	11	5	5	6	10	9	8	8	7	7	5	5	
	4	B	2	5	3	3	6	4	18	11	7	6	10	10	10	7	4	5	5	9	5	5	
	5	B	5	5	4	3	3	4	24	15	11	5	8	9	10	10	6	8	9	5	5	5	
	6	A	5	6	5	3	3	4	27	23	13	5	8	10	8	9	4	7	7	5	5	5	
	7	D	5	5	4	2	2	3	10	14	9	5	8	7	7	6	5	8	5	5	5	5	
	8	C	4	5	4	3	4	4	29	21	13	6	12	12	11	11	6	10	10	6	6	6	
3	1	B	1	2	1	10	18	9	5	4	3	14	12	9	5	10	10	4	8	7	6	7	
	2	A	3	2	1	21	13	7	3	3	2	13	15	10	6	12	9	4	6	6	6	4	
	3	D	3	2	1	12	21	6	2	2	1	13	11	8	5	6	5	4	6	4	4	4	
	4	C	3	2	2	24	23	12	3	4	2	14	16	8	8	11	7	3	5	9	7	7	
	5	C	1	2	1	17	15	9	4	3	2	12	14	9	8	10	6	7	12	9	6	6	
	6	D	0	3	2	12	10	7	3	3	2	11	17	5	5	7	5	5	9	0	5	5	
	7	A	1	1	1	15	15	0	3	3	2	10	11	6	6	9	10	5	8	5	8	5	
	8	B	2	2	1	15	10	10	4	4	3	11	14	8	7	8	8	4	6	6	5	5	
4	1	D	10	8	6	7	7	4	10	9	8	3	4	2	13	13	5	7	8	6	6	6	
	2	C	8	7	5	4	8	5	8	9	7	3	3	3	8	12	4	7	12	9	9	9	
	3	B	0	7	3	5	5	3	7	6	4	1	2	1	7	8	4	11	12	9	9	9	
	4	A	12	10	0	6	6	4	9	10	8	5	6	6	10	10	3	10	13	9	9	9	
	5	A	4	4	4	7	9	4	11	11	8	2	3	1	14	14	0	11	15	9	9	9	
	6	B	9	5	5	3	5	7	7	9	7	3	4	3	7	0	2	7	9	6	6	6	
	7	C	8	9	3	5	7	4	8	11	9	3	4	4	11	13	4	7	10	6	6	6	
	8	D	10	8	4	4	8	5	7	10	0	3	5	5	8	10	4	7	10	6	6	5	

TABLE 3—Continued
TABLE OF INDIVIDUAL RESPONSES AT 48, 72, 96 HOURS
(Data relate to 10 farms, 4 treatment classes at each, 3 cows per class).

Treatment class	Site	Tuber-culin	Cambridge B												Lancashire A																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
			Cow 1						Cow 2						Cow 3						Cow 1						Cow 2						Cow 3																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
			48			72			96			48			72			96			48			72			96			48			72			96																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															

ON A ONE-DIMENSIONAL DIFFUSION METHOD OF ASSAYING ANTIBIOTIC SUBSTANCES AND ITS FUNDAMENTAL FORMULAS

MOTOSABURO MASUYAMA

*Hygieno-meteorological Laboratory, Central Meteorological Observatory,
Tokyo. Institute of Physical Therapy and Internal Medicine, Tokyo Uni-
versity. Institute of Statistical Mathematics, Department of Education.*

1. INTRODUCTION

The cylinder-plate method or the Oxford method for penicillin assay is a direct extension of Fleming's original finding and it is supposed to be more accurate than the ordinary dilution method. However, it needs relatively large amounts of ordinary nutrient agar and of penicillin solution, and it is not suitable, for example, for the estimation of penicillin concentration in serum.

As a quantitative micromethod of assay, the present author and his colleagues, Dr. Torii and Dr. Kawakami, have devised "a one-dimensional diffusion method" or "a small test tube method" which is called in Japan "Zyûsôhô" (literally speaking "a superposition method"). In this paper the author deduces a fundamental formula for this method of assay. The experimental technique and the results will be published in detail elsewhere in Japanese by Torii and Kawakami (1).

2. METHOD

In the original form of the one-dimensional diffusion method a capillary tube was filled with an inoculated agar and its end was inserted in an ampule which contained a solution of penicillin. The test organism was *Staphylococcus aureus* (F.D.A. 209-P). The composition of the ordinary nutrient agar was as follows:

beef infusion	1000cc
NaCl	5g
pepton	10g
agar	15g

(pH = 6.5).

The experimental results showed that only a small amount of the nutrient agar and of the specimen was needed but the technique was not so easy and the growth of bacteria not so good as was expected. To avoid technical difficulties Torii and Kawakami used small test tubes with cotton plugs instead of capillary tubes and ampules. The length and the diameter of the tube is $H = 75\text{mm}$ and $D = 4.5\text{mm}$ ("Murata test tube") or $H = 88\text{mm}$ and $D = 8.0\text{mm}$ ("small test tube") respectively.

To get better results, there are at least three ways, i.e.,

- (i) to use an anaerobe or a facultative anaerobe as the test organism with suitable oxygen donator or growth promoting substances
- (ii) to use a more sensitive strain of bacteria
- (iii) to use a color indicator as is used in the medium of Endo.

In our Institute the following two methods are used

- (i) *Streptococcus hemolyticus*, Murata test tube and medium of following composition devised by Torii:

ordinary nutrient agar	100cc
defibrinated goat blood	10cc
24 hours 1% blood beef broth culture	0.05cc

- (ii) *Staphylococcus aureus*, *Escherichia coli*, *Shigella paradysenteriae* or *Ebertella typhosa*, small test tube and medium of following composition devised by Kawakami:

ordinary nutrient agar	100cc
1% NaNO_3 solution	0.5cc
0.1% methyleneblue solution	3.5cc
24 hours beef broth culture	0.2cc

At first the melted inoculated agar is put into the test tube (approximately 0.5cc for Murata test tube and 2.5cc for small test tube). After the agar has hardened, the solution of antibiotic substance is superposed on it. The amount of solution is approximately one third (for the Murata test tube) or one sixth (for the small test tube) of that of inoculated agar. For ordinary purpose three tubes which contain the same doses are necessary to control random errors of assay which are minimized when the temperature distribution in the incubator is homogeneous. The test tubes are incubated 16 hours at 37°C .

The lengths of inhibition zone are read at least to the nearest 0.5mm. In general there are two frontal surfaces in one tube, i.e., the front of the bacterial growth and that of the hemolysis or the decoloration of the

indicator. We usually read the bacterial front of streptococcus hemolyticus and the front of decoloration of other bacteria. Sometimes the front of decoloration is vague, especially when the tested solution is a culture filtrate. In such a case the deep colored ring in the colored region is used as a front. If we use a more diluted suspension of bacteria and a more concentrated indicator, the front of hemolysis or decoloration, which is formed by the diffusion of the active principles produced by inoculated bacteria, approaches the bacterial front. According to Kawakami's experiments carried out recently it would be better to use 0.02cc of 24 hours beef broth culture and 3.8cc of 0.1% methyleneblue solution. Kawakami's medium should not be exposed in the direct daylight, since the methyleneblue acts as an antibiotic substance in the daylight and the leuco base of methyleneblue is oxidized by the daylight.

All antibiotic substances except Tapeccilline (the commercial name of a sort of culture filtrate of *Penicillium*) used by our colleagues are highly purified, but according to our friends' private communication our method is applicable in the factory for crude extracts with or without slight modifications (low concentration of bacterial suspension, other color indicator).

3. EMPIRICAL FORMULA

Let the concentration of the antibiotic substance and the length of inhibition zone be C and y respectively, then there exists an empirical formula

$$(3.1) \quad y = G[1 - e^{-r(x-a)}],$$

where $x = \log C$.

The following limiting results are apparent:

- (i) $y = G$, for $x = \infty$
- (ii) $\frac{1}{G} \frac{dy}{dx} = r$, for $x = a$,
- (iii) $y = 0$, for $x = a$.

This formula holds well for penicillin and streptococcus hemolyticus or staphylococcus aureus for the range 0.0244 to 200 units per cc. and for patulin, bromsalicil or tapeccilline and other bacteria above cited. The formula is valid for the front of hemolysis or of decoloration or ring; the numerical values of r and G are nearly equal to each other but they are different from the value of a .

TABLE 1
THE LENGTHS OF INHIBITION ZONES IN MM AND THE CONCENTRATIONS
OF PENICILLIN SOLUTION IN U/CC. (FIGURE 1)

			<i>B</i> . . . bacterial front				<i>H</i> . . . front of honolysis			
			10/4		10/4 ²		10/4 ³		10/4 ⁴	
<i>C</i>	10		<i>B</i>	<i>H</i>	<i>B</i>	<i>H</i>	<i>B</i>	<i>H</i>	<i>B</i>	<i>H</i>
<i>B</i>	<i>H</i>									
21.5	20.8		17.9	16.8	14.7	13.8	10.0	9.1	3.5	3.0
21.2	20.1		18.4	17.5	14.8	13.6	9.9	9.6	4.3	3.7
21.8	21.0		18.2	17.2	14.1	13.4	10.3	9.5	4.2	3.5
mean	21.50	20.63	18.17	17.17	14.53	13.60	10.07	9.40	4.00	3.40

To test the goodness of fit, we use the following equation of finite differences. Let w be any given constant, and eliminating the unknown parameter a from the following equation (3.2) and from the previous one (3.1)

$$(3.2) \quad y(x + w) = G[1 - e^{-r(x+w-a)}]$$

we obtain an equation of finite differences

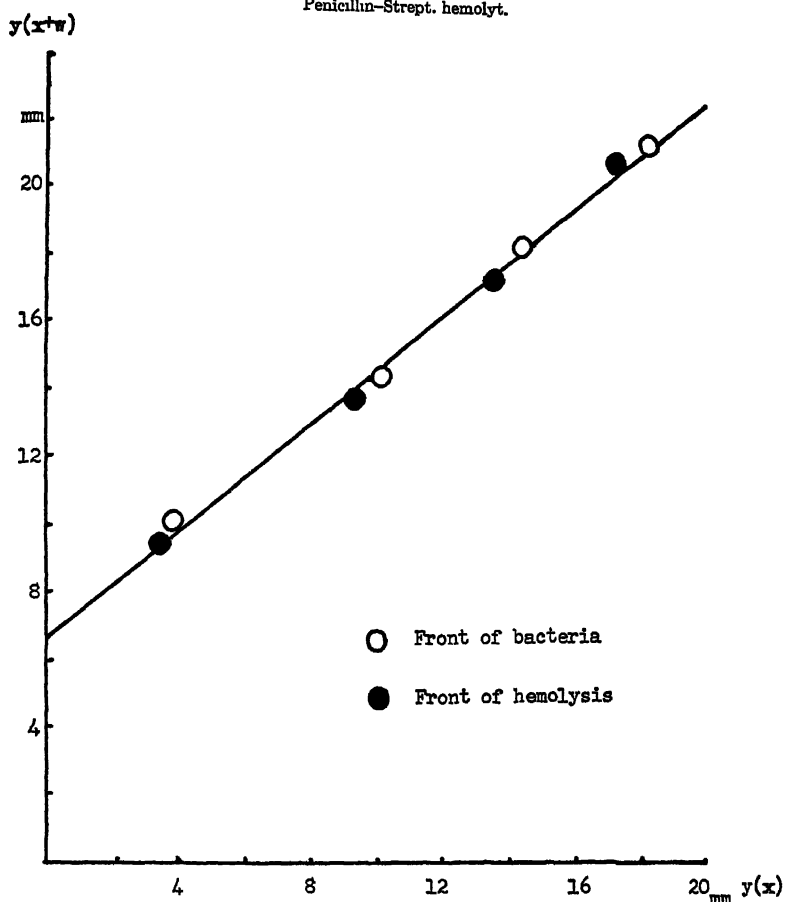
$$(3.3) \quad y(x + w) = e^{-rw}y(x) + G[1 - e^{-rw}]$$

The latter equation shows that though we do not know a priori the numerical values of three parameters in (3.1), we know that there should exist a linear relation between $y(x + w)$ and $y(x)$, i.e., plotting $y(x + w)$ against $y(x)$ we have points on a straight line. From this relation we can estimate at first e^{-rw} or r and then $G[1 - e^{-rw}]$ or G . Substituting these values in (3.1), we can estimate the numerical value of the location parameter a .

To estimate the potency of an unknown solution compared with a standard, the fact that (3.3) is independent of the location parameter a is very important, for the points of the unknown and that of the standard should be on the same line in the $y(x) - y(x + w)$ diagram. To estimate the potency, we need at least

- (i) three standards of different doses S_H , S_M and S_L where the ratios of two successive doses are the same (let it be A) and one unknown U , or
- (ii) two standards of high and low doses S_H and S_L and two unknowns of high and low doses U_H and U_L where both ratios of doses are the same (let them be A).

FIGURE 1. FINITE DIFFERENCES DIAGRAM FOR TABLE 1
 Penicillin-Strept. hemolyt.



Let us use the suffices u and s for the unknown and the standard in the following.

In any assay in which an unknown and a standard are used simultaneously, the potency θ of the unknown is the ratio of the doses of unknown and standard that produce the same response. Consider the log dose response curve of S and that of U . Displace the curve of U by Q to the right so as to place it upon the curve of S , then we have on the one hand

$$(3.4) \quad a_s - a_u = Q$$

and on the other hand

$$(3.5) \quad \log C_u - \log C_s = Q,$$

where

$$(3.6) \quad Q = \log \theta.$$

We shall deduce a formula to estimate the potency in the above cited cases.

I. The diameters are here labeled u_L and u_H for the low and high doses of the unknown, and s_L and s_H for the low and high doses of the standard. The equation of the straight line through two points (u_L, u_H) and (s_L, s_H) on the $y(x) - y(x + w)$ diagram, where $w = \log A$, is

$$(3.7) \quad \begin{aligned} y(x + w) &= \frac{u_H - s_H}{u_L - s_L} y(x) + \frac{u_L s_H - u_H s_L}{u_L - s_L} \\ &= e^{-rw} y(x) + G(1 - e^{-rw}), \end{aligned}$$

and accordingly we have

$$(3.8) \quad e^{-rw} = \frac{u_H - s_H}{u_L - s_L}, \quad \text{and}$$

$$(3.9) \quad G = \frac{u_L s_H - u_H s_L}{u_L - s_L - u_H + s_H}$$

If we transform (3.1) utilizing (3.9), we have

$$(3.10) \quad 1 - \frac{u_L}{G} = \exp [-r(x_{u_L} - a_u)] = \frac{(u_H - u_L)(u_L - s_L)}{u_L s_H - u_H s_L}$$

$$(3.11) \quad 1 - \frac{s_L}{G} = \exp [-r(x_{s_L} - a_s)] = \frac{(u_L - s_L)(s_H - s_L)}{u_L s_H - u_H s_L}$$

Putting $x_{s_L} = x_{u_L}$ we have

$$(3.12) \quad \begin{aligned} &\exp [-r(x_{s_L} - a_s - x_{u_L} + a_u)] \\ &= e^{rQ} = \frac{s_H - s_L}{u_H - u_L} = \left(\frac{u_L - s_L}{u_H - s_H} \right)^{Q/w} \end{aligned}$$

Taking common logarithm of both sides we obtain

$$(3.13) \quad \log \theta = \frac{\log \frac{s_H - s_L}{u_H - u_L}}{\log \frac{s_L - u_L}{s_H - u_H}} \log A$$

Nomograms which facilitate the calculation of the potency by this formula have been made by the author and published by the Japanese

Penicillin Association, Department of Welfare and Public Health. The nomograms consist of two parts, viz., the nomogram for $Z = \log (P/Q)$ and that for $\log \theta = (Z_1/Z_2) \log A$.

II. Let the lengths of inhibition zones for the high, middle, and low dose of S be h , m , and l respectively and that for U be u . Then by the same method of approach we have

$$(3.14) \quad \log \theta = \frac{\log \frac{(G-u)(2m-l-h)}{(m-l)^2}}{\log \frac{h-m}{m-l}} \log A,$$

where θ is the ratio of concentration of low S and U , and we put

$$(3.15) \quad G = \frac{m^2 - lh}{2m - l - h}$$

There is an indeterminate case, where the equation

$$(3.16) \quad 2m - l - h = 0$$

holds by random errors. This case occurs in a narrow range where the log dose response curve could be treated as a straight line. In such a case u may be estimated from this straight line.

4. THEORETICAL FORMULA

Let $u(y, t)$ be the concentration of the antibiotic substance at a point y at a time t and the coefficient of diffusion be D . Then the differential equation for one-dimensional diffusion in porous medium is

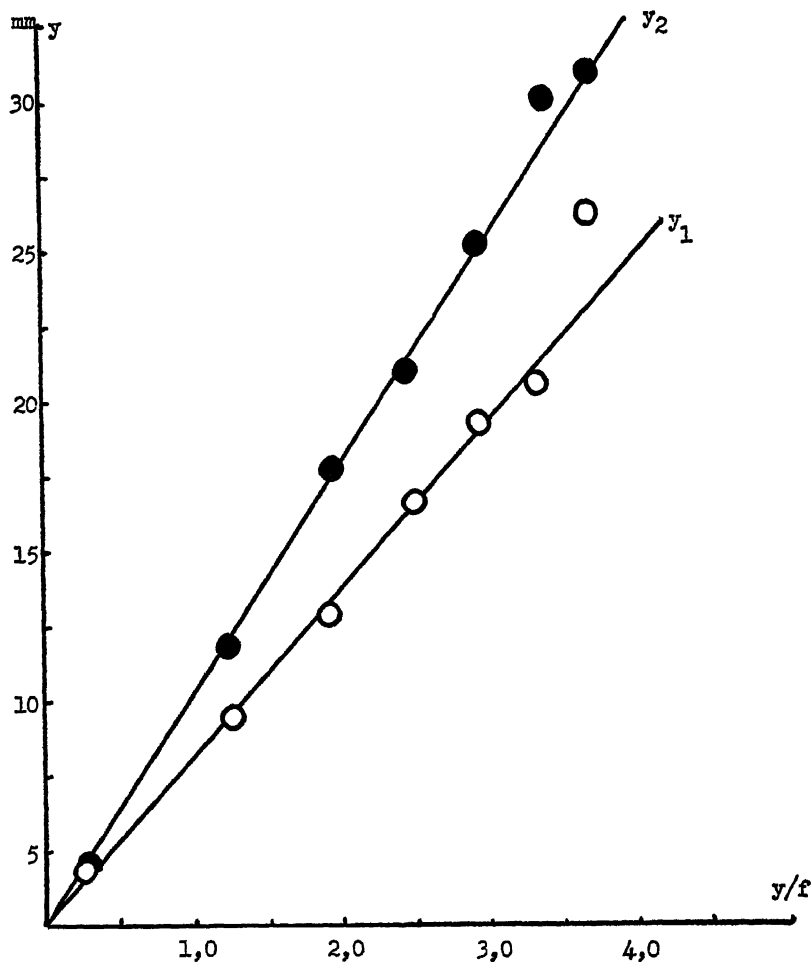
$$(4.1) \quad \frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial y^2}$$

by Darcy's law. The fundamental assumption of the one-dimensionality is based on three experimental facts, i.e.,

- (i) the diameter of the test tube may vary in a certain range to get sufficiently accurate data,
- (ii) the front of hemolysis or decoloration is nearly even, i.e., there is approximately no formation of the meniscus,
- (iii) the minimal effective dose which is estimated by the formula deduced from (4.1) is nearly equal to the mean of the estimates determined by the ordinary dilution method.

If the length K of the inoculated agar column in the test tube is sufficiently large compared with the length of the inhibition zone, we can assume that K is infinite. Let the initial and the boundary condition be

FIGURE 2. GRAPHICAL VERIFICATION OF THE FORMULA (4.5), UTILIZING THE ESTIMATE $k = 0.020$ u/c_s OBTAINED FROM FIGURE 3.



$$(4.2) \quad t = 0; \quad u = 0 \quad \text{for} \quad y > 0, \quad \text{and}$$

$$(4.3) \quad y = 0; \quad u = C \quad \text{for} \quad t > 0$$

Then a well-known solution of the equation (4.1) is

$$(4.4) \quad u = C \left[1 - \sqrt{\frac{2}{\pi}} \int_0^z e^{-t^2/2} dt \right],$$

where $z = y/\sqrt{2Dt}$.

If we let the minimal effective dose and the lag phase be k and τ respectively, then we have as our basic formula

$$(4.5) \quad 1 - \frac{k}{C} = \sqrt{\frac{2}{\pi}} \int_0^{y/f} e^{-t^2/2} dt,$$

where $f = \sqrt{2D\tau}$. To estimate k and f simultaneously by the observed data, we apply Williams' approximation:

$$(4.6) \quad \sqrt{\frac{2}{\pi}} \int_0^y e^{-t^2/2} dt \doteq [1 - e^{-2y^2/\pi}]^{\frac{1}{2}}$$

The relative error of this approximation is for all ranges of y at most 0.7%. Utilizing this sufficiently accurate approximation we have

$$(4.7) \quad -\log \frac{k}{C} \left(2 - \frac{k}{C} \right) = 2y^2/(\pi f^2)$$

When $2C$ is sufficiently large compared with k , we have

$$(4.8) \quad \log C - \log 2k = 2y^2/(\pi f^2) = y^2/\pi D\tau$$

This equation indicates that if y^2 is plotted against $x = \log C$, the plotted points will be on a straight line. The point of intersection of this line and the x or horizontal axis and the slope of the line give the numerical value of $\log 2k$ or k and that of $2/(\pi f^2)$ or f respectively.

Utilizing this theoretical formula we can estimate how the length of the inhibition zone increases when the test tube has been stored in a refrigerator before incubation, because D in f varies proportionally to the absolute temperature (approximately) and τ increases by duration of refrigeration. As the biological or physical meanings of the three parameters k , D , and τ are clear in this case, theoretically speaking, (4.8) is better than (3.1) if the former fits the actual data. Formally speaking, in (3.1) $y = 0$ corresponds to the minimal effective dose, or in other words, there seems to exist the relation

$$(4.9) \quad k = e^a$$

but we have no evidence which indicates the validity of the empirical formula for sufficiently small values of y . The uncertainty of such an induction may be easily seen in (4.8), for if we take (4.8) as an empirical formula, $y = 0$ corresponds to $C = 2k$, i.e., the estimated minimal effective dose is twice as large as the true one.

To test whether or to what extent this formula holds, it is desirable to dilute the solution in question in geometric progression and then the x are in arithmetic progression and accordingly the corresponding y^2

FIGURE 3. $\log C - y^2$ DIAGRAM

Penicillin-Strept. hemolyt.

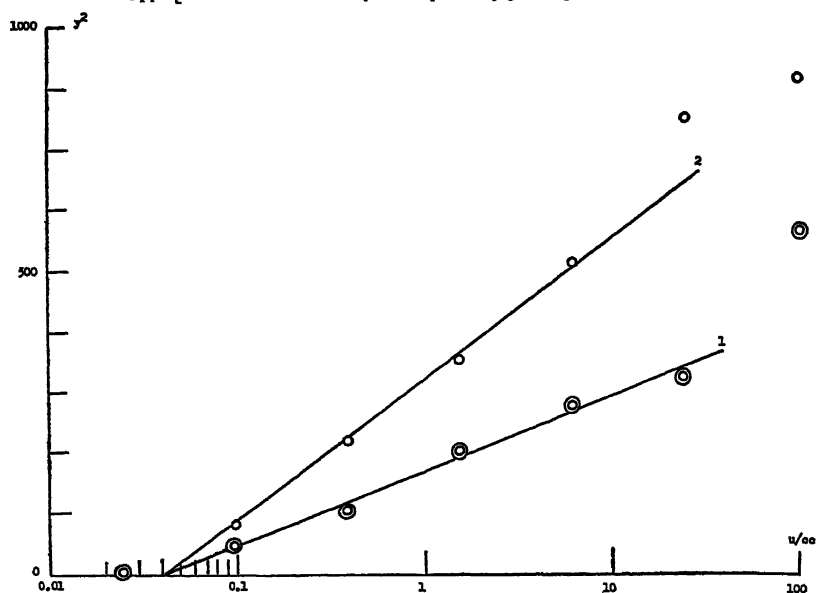
Upper and lower curve correspond respectively y_2 and y_1 of the Table 2.

TABLE 2

THE MEAN LENGTHS OF INHIBITION ZONES IN MM AND CONCENTRATIONS OF PENICILLIN SOLUTION IN U/CC. (FIGURE 2 AND 3)

y_1 . . . the solution is immediately superposed

y_2 . . . the solution is superposed after 24 hours refrigeration

C	y_1	y_2	y/f
100	23.9	28.6	3.72
100/4	18.1	27.5	3.35
100/4 ²	16.7	22.7	2.95
100/4 ³	14.2	18.7	2.49
100/4 ⁴	10.4	14.9	1.95
100/4 ⁵	6.97	9.23	1.27
100/4 ⁶	1.77	1.97	0.23

are also in arithmetic progression. As the actual tube has finite length, there is a systematic deviation for large values of y , especially when we use Murata test tubes. In such a case we should use the longer test tubes.

This formula is valid for penicillin and streptococcus hemolyticus or staphylococcus aureus, for patulin and staphylococcus aureus, eber-

tella typhosa or shigella paradysenteriae and for tapecolline and the various bacteria cited above.

To estimate the potency of an unknown solution U absolutely by the formula (4.8), dilute the solution U in geometric progression with a common ratio A , then the first differences of y^2 should be constant within a certain range where the assumptions hold. Let the concentration of the original solution be W , and then that of the $(n + 1)$ st solution is $C_n = WA^{-n}$. Now the equation (4.8) gives

$$(4.10) \quad \log C_n - \log 2k = \log W - \log 2k - n \log A = 2y_n^2/\pi f^2$$

where y_n means the length of the inhibition zone for C_n . Plotting y_n^2 against n , we can estimate $\log W - \log 2k$ or W/k , i.e., the multiple of the minimal effective dose.

According to the experimental data which were made by my colleagues using sodium penicillin G, the numerical values of k estimated by (4.8) for staphylococcus aureus (F.D.A. 209-P) lay in a narrow range which contained $k = 0.02$ units per cc; the estimates based on the means of readings of three test tubes, distributed in a range from 0.015 to 0.030 units per cc. The estimated values of k were approximately constant within a series of experiments done simultaneously, but were slightly different from day to day, even though each set of data agreed very well with the theoretical curve. The reason for such a fluctuation is not clear at present. According to our theoretical formula the concentration of agar or the variation of lag phase due to various thermal conditions might vary the numerical value of f but should not vary that of k . There remain at least two possibilities:

- (i) k may vary under various conditions
- (ii) the front may move after its formation.

To estimate k absolutely, the front of decoloration may not be suitable, for it is formed by the diffusing active principles. In routine work, it would be convenient to use semi-logarithmic paper in which a square scale is taken along the vertical axis. Then research workers may learn the potency without any calculation by plotting y against C .

To estimate the potency of an unknown solution U compared with a standard solution S , we can use the well-known formula

$$(4.11) \quad \log \theta = \frac{U_H - S_H + U_L - S_L}{S_H - S_L + U_H - U_L} \log A,$$

where the squares of the measured diameters are labeled U_L and U_H for the low and high doses of the unknown, and S_L and S_H for the low and high doses of the standard.

Finally we want to consider the meaning of the empirical formula, compared with the theoretical one. Transforming our empirical formula (3.1), we have

$$(4.12) \quad y = G[1 - \exp \{-r(\log C - a)\}] = G[1 - C^{-r}e^{ar}],$$

and accordingly

$$(4.13) \quad 1 - \frac{y}{G} = C^{-r}e^{ar} = \left(\frac{k}{C}\right)^r,$$

where we put $\log k = a$, assuming the validity of (4.9). Comparing (4.5) with (4.13) we can conclude that the empirical formula is based on a parabolic approximation of the probability integral, i.e.,

$$(4.14) \quad \sqrt{\frac{2}{\pi}} \int_{y/f}^{\infty} e^{-t^2/2} dt \doteq \left(1 - \frac{y}{G}\right)^{1/r}$$

To know the order of approximation we let the left side of (4.14) be $p(y/f)$; then if both sides of (4.14) are exactly equal to each other, the equation

$$(4.15) \quad p^r(t) = 1 - \frac{f}{G} t, \quad t = y/f$$

should hold for every value of t . Transforming (4.15) we have

$$(4.16) \quad t = \frac{G}{f} [1 - e^{r \log p(t)}],$$

which has the same functional form as (3.1), and we can test the validity of the equation (4.16) by the calculus of finite differences.

The exact solution of the differential equation (4.1) under the more plausible initial and boundary conditions,

$$(4.17) \quad \begin{array}{lll} t = 0 & : & u = 0, \\ y = 0 & : & u = Ce^{-bt} \quad (b > 0), \\ y = K & : & \partial u / \partial y = 0, \end{array}$$

has been obtained in the form of a Fourier series. However, we can hardly utilize this form of solution, because the unknown parameters are included in each term of the series.

5. CONCLUSION

The author and his colleagues, Dr. Torii and Dr. Kawakami, have picked up the essential part of the cylinder plate method of assaying penicillin and devised a new method of assaying antibiotic substances.

The basic idea is to utilize the one-dimensional diffusion with suitable indicators. The empirical formula (3.1) and the theoretical one (4.8) fit very well the data obtained by Torii, Kawakami, and Kozima. Even though formulas (3.1) and (4.5) hold for a wider range, it is impossible to estimate the potency without any standard. To estimate the potency without any standard, formula (4.8) is useful, provided that the inequality $2C > k$ holds and that K is sufficiently large. Our superposition method of the absolute measurement of the potency seems to be better than the ordinary dilution method. In the dilution method only two successive test tubes are used to estimate the potency and other test tubes remain unused. Furthermore, it has one serious defect, i.e., in the dilution method we must determine the point of contact of the dose response curve on the dose axis. It is well-known that this is very difficult and inaccurate. In our method we use all sufficiently accurate readings of all the test tubes without turbidimeter or colorimeter to estimate the potency, and we can estimate the error of estimated potency by utilizing the analysis of covariance. However, this method of assaying antibiotic substances might not be suitable for the electro-positive substances, for the latter are adsorbed by the electro-negative agar and the length of the inhibition zone is too short to estimate its potency. In such a case it would be desirable to use their neutral or electro-negative double salts.

It is worth noting that with a slight modification of the interpretation of the parameters the formulas might be applicable in wider field of assaying chemicals and biological products. In fact, the author and Dr. Okawara have applied formula (4.8) to estimate the potency of the pepsin solution, using the edestin-sodium chloride agar and the pepsin solution in place of the inoculated agar and the penicillin solution. In this case the length of the digestion zone y increases proportionally to the square root of the time of reading τ , (at least within several hours), which is expected naturally from (4.8).

The theory of errors in the estimation of potency will be developed in the near future, being based on the observed data. At present to estimate the potency absolutely the analysis of covariance method is applicable. The classical method of approach, i.e., the theory of large samples as is used in Knudsen & Randall's paper (2) might not be suitable for small samples used in routine work.

REFERENCES

- (1) Torii, Kawakami & Kozima. On a One-dimensional Diffusion Method of Assaying Penicillin and other Antibiotic Substances. *Jour. Japanese Penicillin Assoc.* 1, 281, 1947.
- (2) Knudsen & Randall. Penicillin Assay and its Control Chart Analysis. *Jour. Bact.* 50, 187, 1945.

ROUTINE COMPUTATION OF BIOLOGICAL ASSAYS INVOLVING A QUANTITATIVE RESPONSE

M. J. R. HEALY

Rothamsted Experimental Station

SUMMARY

A NOMOGRAM is described for rapid computation of routine biological assays using a 6-point design.

INTRODUCTION

Many biological assays involving a quantitative response are based on a straight line relationship between response and the logarithm of the dose.

A common design for such assays is the 4-point design, in which two doses of the unknown preparation are compared with two doses of a standard. Several writers have discussed the routine computation of such assays, and have constructed nomograms for the purpose (Knudsen, 1945; Knudsen and Randall, 1945; Bliss, 1946).

In a general survey of assay problems, Finney (1947) has pointed out that the 4-point design does not give a test of linearity of the dose-response curve, and hence of the assumptions on which the assay is based; he recommends the use of at least three levels of each preparation, giving rise to a 6-point assay. When a large number of these assays are being carried out, it is often possible to standardise the experimental procedure so that the spacing between doses and the number of replicates at each dose remain constant; in this case the computation is much reduced by the use of the nomogram described below.

COMPUTING PROCEDURE

The doses of both preparations should be related by a constant dilution factor I , giving equal spacing on the log. scale. It is supposed that n readings are taken at each dose.

TABLE 1
COMPUTING SHEET FOR 6-POINT ASSAY

	Standard			Unknown		
	0	1	2	0	1	2
	48	60	84	56	76	92
	40	76	84	56	76	90
	46	62	84	44	77	88
Sum	$S_0 = 134$	$S_1 = 198$	$S_2 = 252$	$U_0 = 156$	$U_1 = 229$	$U_2 = 270$
Range	8	16	0	12	1	4

Total Range $R = 41$ Dilution Factor $I = 1.5$

$$S_2 + S_1 + S_0 = +584$$

$$U_2 + U_1 + U_0 = +655$$

$$S_2 - S_0 = +118$$

$$U_2 - U_0 = +114$$

$$S_2 - 2S_1 + S_0 = -10$$

$$U_2 - 2U_1 + U_0 = -32$$

$$\text{Materials } D = (U_2 + U_1 + U_0) - (S_2 + S_1 + S_0) = +71$$

$$\text{Slope } B = (U_2 - U_0) + (S_2 - S_0) = +232$$

$$\text{Parallelism } P = (U_2 - U_0) - (S_2 - S_0) = -4$$

$$\text{Curvature i } H = (U_2 - 2U_1 + U_0) + (S_2 - 2S_1 + S_0) = -42$$

$$\text{Curvature ii } K = (U_2 - 2U_1 + U_0) - (S_2 - 2S_1 + S_0) = -22$$

$$t_1 R = 29.9$$

$$t_2 R = 51.7 \text{ (5\% level)}$$

$$D/B = +0.306$$

$$D/R = +1.73$$

$$\text{Relative potency} = 1.087 - 1.180 - 1.294 \text{ (5\% limits)}$$

A suitable computing sheet is shown in Table 1. The responses are entered on the sheet and their sums and ranges are found. The ranges are totalled to give a value R and the sums are combined, using "factorial coefficients" (Emmens, 1948, p. 92), as described on the sheet. For the assay to be considered a valid one at a given level of significance, the last three totals must not numerically exceed certain values; the limit for P is $t_1 R$, and the limit for H and K is $t_2 R$ where t_1 , t_2 are factors tabulated, Table 2, for significance levels of 5% and 1%.

Provided the assay proves to be valid, the estimate of relative potency is found from the nomogram shown in Fig. 1. The quantities D/B and D/R are calculated and the corresponding points marked off on the scales AA' and BB' . The points are joined, and the relative potency with its limits of error can be read off from the scales on OO' , PP' and QQ' .

In this scheme, the precision of the assay is determined by the use of range in place of standard deviation, in order to avoid computing sums of squares. The resulting estimate of error is not the best possible

TABLE 2
FACTORS FOR ASSESSING THE VALIDITY OF THE ASSAY

<i>n</i>	<i>t</i> ₁		<i>t</i> ₂	
	5%	1%	5%	1%
2	0.994	1.520	1.72	2.63
3	0.730	1.034	1.26	1.79
4	0.674	0.926	1.17	1.61
5	0.657	0.894	1.14	1.55
6	0.654	0.883	1.13	1.53
7	0.659	0.887	1.14	1.54
8	0.666	0.894	1.15	1.55
9	0.677	0.902	1.17	1.56
10	0.685	0.914	1.19	1.58

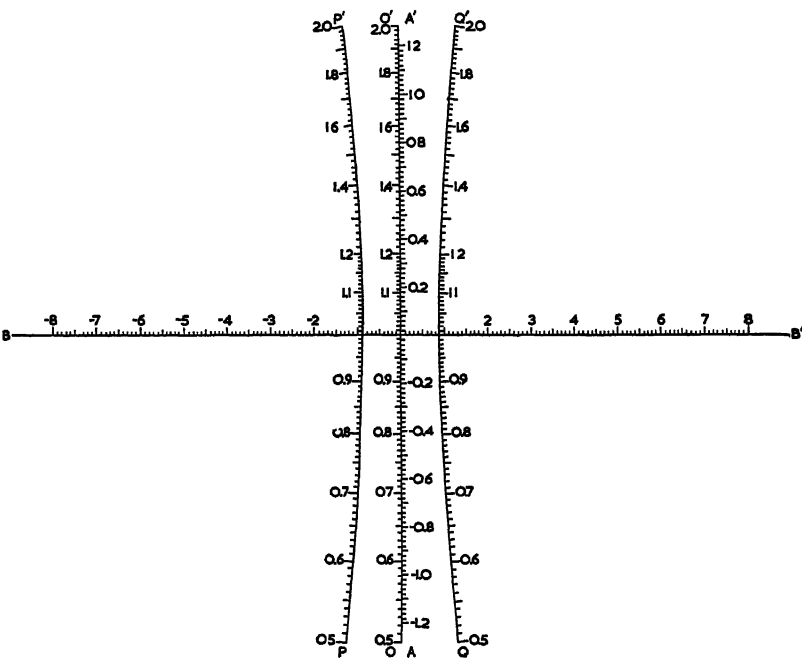


FIGURE 1.
NOMOGRAM FOR 6-POINT ASSAY

$n = 3; \quad I = 1.3; \quad 5\% \text{ limits.}$

LOCATE D/B ON AA' AND D/R ON BB' . JOIN, AND READ RELATIVE POTENCY ON OO' AND LIMITS OF ERROR ON PP', QQ' .

one except when $n = 2$; in addition, it may be biased if the underlying law of variation is not normal. These drawbacks must be set against the added convenience of the use of range. They are unlikely to be of practical importance when n is small, as will usually be the case.

NUMERICAL EXAMPLE

Table 1 gives the results of an assay of nisin, an antibiotic substance derived from *Str. lactis* (Mattick and Hirsch, 1947). A strain of the acid-forming bacteria, *Str. agalactiae* is grown in the presence of a sub-lethal quantity of the antibiotic under controlled conditions, and it is found that the amount of acid produced after about 12 hours is related to the quantity of nisin. The exact form of the relationship is complicated, but a useful linear range can be obtained by plotting pH against log. dose, and this fact has been used in the assay under consideration. A standard preparation containing 50 units/ml. of nisin was used as dose S_2 , and diluted twice in the ratio 1.5 : 1 to give doses S_1 and S_0 . Similarly, the unknown preparation and two successive dilutions from it were used for doses U_2 , U_1 and U_0 . The responses (which are given in terms of $100 \times (\text{pH} - 5)$), are combined as described above to give $D/B = + 0.306$, $D/R = + 1.73$. The assay is shown to be valid by comparing the values of P , H and K with the limits of t_1R and t_2R , and the use of the nomogram gives the relative potency as 1.180, with 5% limits of 1.087 to 1.294. Thus, as the standard contains 50 units of nisin per ml., the unknown is estimated to contain 59.0 units/ml. with 5% limits of 54.4 to 64.7 units/ml. I am indebted to Dr. A. Hirsch of the National Institute for Research in Dairy for permission to use these figures.

CONSTRUCTION OF THE NOMOGRAM

The ordinary theory of biological assay (see, for example Emmens, 1948) shows that the estimate of relative potency is given by $\log I \times 4D/3B$. If we write $\alpha = \sqrt{2}D/\sqrt{3}B$, it can be shown that the fiducial limits of α are given by the roots of the quadratic equation

$$(\sqrt{2}D - \sqrt{3}B\alpha)^2 = 12nt^2s^2(1 + \alpha^2)$$

where s^2 is the error variance of a single response and t is the appropriate deviate of the ordinary Student distribution. In the present application, t is replaced by an analogous quantity which allows for the replacement of s by the sample range (Lord, 1947). The limits of the estimate of relative potency are then obtained by a suitable multiplying factor.

The nomogram in Fig. 1 is a graphical representation of these relations, and will be described in terms of rectangular coordinates x and y . The axes are drawn first, and suitable units of x and y chosen—in practice the unit of y may conveniently be ten times the unit of x . On the x -axis, a linear scale BB' is marked off, each unit of which has a length depending on n and the level of significance used. This unit is tabulated below.

TABLE 3
LENGTH OF UNIT DIVISION ON SCALE BB'

n	2	3	4	5	6	7	8	9	10
5%	0.821	1.119	1.212	1.243	1.248	1.240	1.227	1.206	1.192
1%	0.537	0.790	0.882	0.913	0.925	0.921	0.913	0.905	0.893

Two scales are marked on the y -axis; the first of these is another linear scale AA' , each unit having a length of $\sqrt{(2/3)} = 0.8165$. The second scale OO' is logarithmic, and the position of the graduation corresponding to each potency ratio can be found from the relation $y = 0.6124 \times \log \text{potency ratio} / \log I$. The two curves PP' , QQ' are given by the equation $x^2 = 1 + y^2$, and the scales on them are graduated by the same logarithmic relation. The example shown in Fig. 1 is for use when $n = 3$, $I = 1.5$ and the 5% limits of error are required.

REFERENCES

- Bliss, C. I. (1946). *A Revised Cylinder-Plate Assay of Penicillin*. J. Amer. Pharm. Assoc., 35, 6-12.
- Emmens, C. W. (1948). *Principles of Biological Assay*. London: Chapman and Hall, Ltd. 1st Edition.
- Finney, D. J. (1947). *The Principles of Biological Assay*. Suppl. J. Roy. Stat. Soc., 9, 46-91.
- Knudsen, L. F. (1945). *Penicillin Assay*. Science, 101, 46-48.
- Knudsen, L. F. and Randall, W. A. (1945). *Penicillin Assay and its Control Chart Analysis*. J. Bacteriology, 50, 187-200.
- Lord, E. (1947). *The Use of Range in place of Standard Deviation in the t-test*. Biometrika, 34, 41-67.
- Mattick, A. T. R. and Hirsch, A. (1947). *Further Observations on an Inhibitory Substance (Nisin) from Lactic Streptococci*. Lancet, 253 (2), 5-8.

QUERIES

QUERY: In the course of running a series of experiments, I
73 have encountered a problem which is probably simple to solve but which I am incapable of solving. The problem is this:

Given two distributions with sample means \bar{x} and \bar{y} , and corresponding estimated standard errors, s_x and s_y , what is the reliability of the observed ratio, \bar{x}/\bar{y} , and how does one go about the setting of confidence limits?

It seems to me that the problem I have stated is one of great practical significance—it would not otherwise have been called to your attention.

ANSWER: Your problem is certainly of practical importance. In biological assay, for example, chief interest lies usually in the ratio of two weighted means of responses, and not in differences of means. If we write the ratio of your two unweighted means as

$$m = \frac{\bar{x}}{\bar{y}},$$

the method commonly used is to take

$$V(m) = \frac{1}{\bar{y}^2} \{ V(\bar{x}) + m^2 V(\bar{y}) \}, \quad (1)$$

and to calculate confidence limits with the aid of this variance formula. This must be condemned, since it seriously overestimates the precision of m except when \bar{y} is very much larger than its standard error; only if \bar{y}^2 is at least 40 $V(\bar{y})$ for 95% limits, or 70 $V(\bar{y})$ for 99% limits, can equation (1) be safely used.

A better procedure is based upon a theorem first stated by Fieller in 1940 (Journal of the Royal Statistical Society, Supplement, Vol. 7, pp. 1-64). Suppose that \bar{x} , \bar{y} are means of n_1 , n_2 observations respectively, from distributions that may be assumed normal. Suppose further that the two distributions have a common variance, which is estimated by a mean square, s^2 , with f degrees of freedom (f may be

$n_1 + n_2 - 2$, but can differ from this if the means \bar{x} , \bar{y} are from an analysis of variance and not just from two simple samples). Then, for any value of a quantity μ , $(\bar{x} - \mu\bar{y})$ is normally distributed,

$$V(\bar{x} - \mu\bar{y}) = s^2 \left(\frac{1}{n_1} + \frac{\mu^2}{n_2} \right),$$

and therefore

$$t = \frac{\bar{x} - \mu\bar{y}}{s \left(\frac{1}{n_1} + \frac{\mu^2}{n_2} \right)^{1/2}} \quad (2)$$

follows the t -distribution with f degrees of freedom. If we now determine μ by the condition that t shall be the tabular entry for an agreed percentage point, the two values of μ must be the lowest and the highest which are not significantly in conflict with

$$E(\bar{x}) - \mu E(\bar{y}) = 0;$$

thus they are fiducial limits to m . The equation for these limits is a quadratic, and the two roots, m_L , m_U , may be found directly by the formula

$$m_L, m_U = \left\{ m \pm \frac{st}{\bar{y}} \left(\frac{1 - g}{n_1} + \frac{m^2}{n_2} \right)^{1/2} \right\} \div (1 - g), \quad (3)$$

where

$$g = \frac{s^2 t^2}{n_2 \bar{y}^2};$$

of the alternative signs in (3), “−” will give the lower, “+” the upper fiducial limit. Note that when g is negligibly small the formula is the same as that based upon (1), but when g is, say, greater than 0.1, equation (1) may be seriously misleading. The limits given by (3) are true *fiducial* limits to the ratio; I understand that doubts have been expressed as to whether they are also *confidence* limits.

Fieller's theorem will still apply when \bar{x} , \bar{y} are means of n correlated variables, or when they are weighted means, either correlated or not; weighted means include, for example, regression coefficients on another variate. Fisher gives the method for correlated means in Section 62.1 of his *Design of Experiments*, and a generalization of (3) which includes many cases appears as formula (4.7) in my *Probit Analysis*. You may wish to know what happens if the variances in the populations from

which \bar{x} , \bar{y} come are not equal. Separate mean squares, s_1^2 and s_2^2 , would then have to be used, and equation (2) would be modified so as to replace t by a deviate from the Fisher-Behrens distribution (see Fisher and Yates's *Statistical Tables*, Table V_1). An explicit formula such as (3) cannot be given, and the limits must now be obtained by using a method of successive approximation to solve equation (2) for μ .

D. J. FINNEY

QUERY: We would very much appreciate your advice regarding application of the t -test to accumulated chi-square values for a time series investigation.

74 The samples are frozen peas stored at -10°F and 0°F for periods of 4 to 40 weeks. Judges evaluated their quality at about 4 week intervals. The -10°F and 0°F samples and a duplicate of one of these samples were submitted each judging period (about 12 judges, 5 replicate judging periods for each storage period). Judges were asked (1) to indicate whether there was any difference between samples, (2) if there was a difference to check duplicates, and (3) if there was a difference in flavor, texture and color to indicate which sample or samples were best. Data were analyzed by chi-square. In the case of best flavor, texture, color data only the results for those who identified duplicates were included in the chi-square analyses.

The results for identification of duplicates seem inconsistent. Chi-square values were significant at 4, 14, 18, 20, 32, 36 and 40 weeks, almost significant at 24 weeks but definitely not significant at 8 and 28 weeks (see condensed summary in Table 1 below).

The t -test applied to chi-square values accumulated up to each storage period has been suggested for these data:

$$t = \frac{S\sqrt{\chi^2}}{\sqrt{N}}$$

R. A. Fisher recommended this test to M. P. Masure of this laboratory in correspondence in 1931 for data similar to ours. Masure used the test in his publication, Effect of Ultraviolet Radiation on Growth and Respiration Pea Seeds, with Notes on Statistics, *The Botanical Gazette* 93, 21-41, 1932.

The chi-square formula we used was

$$\frac{(| \text{observed-expected} | - 0.5)^2}{\text{expected}} .$$

Expected, of course, was $1/3$ total n for identified and $2/3$ total n for failed to identify. We permitted judges to say that there was no difference between the 3 samples and added $1/3$ "no difference" judgements to identified and $2/3$ to failed to identify (judges who said there was a difference but checked the wrong samples as duplicates). Some investigators insist that judges always choose two samples as duplicates. We think it psychologically wrong to force a choice when the judge has done his best and still cannot detect any difference between samples. If there is any statistical advantage for forcing a choice or permitting "no difference" answers, we think the advantage in favor of the latter because the number who indicate no difference is small and might not distribute normally.

Storage Period Wks.	Number of Judgements	Identifications		No Difference
		Correct	Incorrect	
4	51	21	16	14
8	52	12	28	12
12	75	30	28	17
16	68	39	23	6
20	68	34	29	5
24	61	26	30	5
28	68	22	38	8
32	76	42	30	4
36	68	40	28	0
40	60	36	24	0

If we have made a suitable application of the t -test, we think we can conclude that the difference between samples is significant at 18 weeks. Our queries are: have we applied a proper statistic and is our conclusion correct for these samples? Also, do our data show anything else of statistical significance? Do you know of a literature reference to the application of the t -test as discussed here? Do you suggest some other treatment of our data?

We must be careful to separate two questions:—"How soon have I accumulated evidence that there is a real difference at a given level of significance?" and "How does the difference seem to change with time?" There are points worthy of careful attention in both cases.

The process of continually trying combining the present trial with

those already accumulated, testing the significance of all the data to date, and continuing until "significance" is reached can be dangerous. If really carried out in this way, it is sure to reach "significance" no matter what the true state of affairs. You are exempt from the worst features of this difficulty, because you have previous knowledge that the two storage temperatures will really be different after a long enough time. (However, others might find real difficulty here.)

The method outlined by Masure is surely sound, provided, as he carefully stated, that all the deviations involved in the chi-squares were in the same direction. However, when you are calculating chi-squares to be *combined*, you should not make the 0.5 correction for continuity, as Cochran (Iowa State College Jour. Sci. 16:421-436, 1942) has shown.

A simpler way to analyze data such as yours, where the null hypothesis of pure chance gives a *fixed* chance of one-third of identifying correctly, is the following: Accumulate the number of correct identifications, accumulate separately the number of incorrect identifications, and test to see if the ratio is significantly different from 1 to 2. Thus, after 20 weeks, you have a total of 136 correct and 124 incorrect (and 54 "no difference" which we discard for the present). We can test this against the 1 to 2 ratio using

- (i) chi-square (where we do make the 0.5 correction)

$$\chi^2 = \frac{(136 - 86.7 - 0.5)^2}{86.7} + \frac{(124 - 173.3 + 0.5)^2}{173.3} = 59.8$$

which is very significant on one degree of freedom,

- (ii) a simple graphical method (see Frederick Mosteller and John W. Tukey, "The uses and usefulness of binomial probability paper", *Jour. Amer. Stat. Assn.*, 1949, Example 1)
- (iii) the simple formula for an approximate normal deviate corresponding to the use of binomial probability paper

$$\sqrt{136 \left(\frac{2}{3} \right)} - \sqrt{(124 + 1) \frac{1}{3}} = 6.13,$$

where we have multiplied each observed number by the *other* theoretical probability and have increased that observed number by unity which *reduces* the difference of the square roots. A normal deviate of 6.13 is also very significant. This last formula gives as accurate results as the corrected chi-square and is very convenient with a slide rule.

You included the judgements of "no difference" in your analysis. For

the purpose of judging significance alone, there seems to be no reason to include them, and several to leave them out. But if you want an indication of the magnitude of the differences between storage temperatures, *as measured by the ability of these judges to detect differences correctly*, then these judgements of "no difference" are important, and should come in. It seems natural to score as follows:

correct identification	+1
incorrect identification	-1/2
no difference	0

and then to take the average score, which we may express as a percentage for convenience, as a measure of difference. Arithmetically, this simplifies to

$$\frac{(\text{correct}) - 1/2(\text{incorrect})}{(\text{total judgements})}$$

and should have a variance due to sampling of not more than

$$\frac{1}{2(\text{total judgements})}.$$

Your ten trials score 25%, -4%, 21%, 40%, 29%, 18%, 4%, 35%, 38%, 40%. The greatest standard deviation expected from sampling is about

$$\frac{1}{\sqrt{2(65)}} = .088,$$

or about 8.8%, and a value somewhat smaller than this is reasonable. With the exception of the values at 8 and 28 weeks, your results seem to be consistent with a true score of about 30, *not* changing with time. Could you make a test after one day's storage?

JOHN W. TUKEY

THE BIOMETRIC SOCIETY

Benelux Meeting. Following some admirable preliminary work by our Council member, Dr. Nuerdenburg, the State University at Utrecht was host for a program of papers on biometry this last July 2nd. Dr. W. A. Mijsberg, Professor of anatomy and embryology at Utrecht, discussed frequency curves of stature from 1811 to the present time with particular relation to the effect of economic status upon the ability to reach full genotypical height. Director P. deWolff of the Municipal Statistical Bureau of Amsterdam discussed statistical problems concerned in the assay of vitamin D_3 in chickens. Mr. J. A. Enters, an industrial statistician, reported on the use of measurements of Dutch men and women for reducing fitting costs in the clothing industry. Dr. E. van der Laan of the Agricultural State College at Wageningen surveyed the different types of experimental design which are now available. Dr. G. A. Gussenhoven outlined his results in combining factors recorded in the case histories of patients with pulmonary tuberculosis. Dr. R. A. M. Bergman, Professor of the Medical Faculty of Batavia, described his researches on the linear growth of snakes. Dr. J. ten Doesschate reported on some quantitative aspects of the gerontology of the eye.

About 30 scientists attended the meeting, including one from Belgium, and at the close of the all-day session the audience voted for a second meeting. Another result has been the material increase in the membership of the Society in Holland. Dr. Neurdenburg is to be congratulated on so auspicious a beginning toward the development of a Benelux Region.

The Second International Biometric Conference. The Second International Biometric Conference convened at the University of Geneva in Switzerland on August 30, and continued for four days with a total registration of 102. The Governments of Belgium, Great Britain, Greece, Italy, Netherlands, Portugal, Spain and Venezuela named official delegates. Thirteen others attended as delegates of Academies of Science, international organizations, municipalities, Regions of the Biometric Society or other organizations. Some 19 countries were represented in all, Great Britain leading with 18, and followed by Switzerland with 14,

France 10, Italy 10, Netherlands 10, Denmark 9, United States 9, Belgium 5, India 3, Portugal 3, Argentina 2, Australia 2, and Canada, Finland, Greece, Mexico, Spain, Sweden and Venezuela with one each.

The Conference opened with a welcoming address by Professor G. Tiercy, Rector of the University of Geneva, and by Professor A. Franceschetti of the Faculty of Medicine, who spoke successively in French, English and Italian, and continued with a business meeting of the Society. The afternoon session on experimental design with Dr. Yates in the chair, featured papers by Professors Cox and Quenouille, with discussions by Drs. Hald, Åstrand, Rasch, Bernstein, Schutzemberger, Fisher, Healy and Bartlett, and a concluding summary by the chairman. At the end of the session Professor and Mrs. Franceschetti entertained members of the Conference with a delicious high tea at their country home on the shores of Lake Geneva where we were also rewarded with a splendid view of Mt. Blanc in the setting sun.

The morning session on August 31 concerned the recent applications of biometrical methods in genetics under the chairmanship of Professor Fisher. The papers by Drs. Yates, Cavalli and Finney were discussed individually, both by the speakers and in addition by Profs. Cochran, Pompilj, Chodat, Haldane, Bernstein and Healy. The afternoon session on biometrical aspects of biological assay under the chairmanship of Dr. Finney offered papers by Drs. Irwin and Perry and a lively discussion by Drs. Bliss, Jerne, Tripod, Fieller, Hartley, Bernstein, Rasch, Martin and the chairman.

On September 1 the morning session concerned the present status of biometry with Professor Darmais in the chair and a remarkably lucid paper by Prof. Cochran which was discussed by Drs. Hopkins, Mahalanobis, Gini, Haldane, Rasch, Kemp and Rapaport. The afternoon session on industrial applications of biometry was chaired by Dr. Åstrand and had as its principal speaker Dr. Davies with discussions by Miss Day, Drs. Fieller and Hald. That evening the members of the Biometric Conference and of the International Union for the Study of Populations, which was meeting concurrently in Geneva, were entertained at a reception by the State Council of the Canton of Geneva and by the Municipal Council of the Town of Geneva at Palais Eynard, a most colorful affair.

The session of the morning of September 2 under the chairmanship of Prof. Mahalanobis considered teaching and education in biometry. The principal paper by Professor Bartlett was followed by an active discussion which included Profs. Darmais, Cochran, Cox, Gini, Roy and Vessereau, also Drs. Finney, Bliss, Yates, van der Laan and Martin. At the conclusion of the session it was resolved that teaching material

on Biometry be assembled by the Society for distribution to members and others who are interested. The final scientific session on the afternoon of September 2 under the chairmanship of Dr. Buzzati-Traverso consisted of four contributed papers read by Drs. Rapaport, Boeri, Schwartz and Nass. A final business meeting concluded the Conference.

Most of the principal papers had been mimeographed in advance so that copies were available for those attending. English and French were official languages and all discussions were translated most competently from one language into the other. A photographer was active during the conference and the morning session on August 31 was interrupted for a group photograph on the steps of the University building. It is planned to publish the Proceedings of the Conference as completely as possible in *BIOMETRICS* during the coming year with the aid of a grant of \$800. from UNESCO. It is hoped that reprints will be available of any papers which are published elsewhere for distribution to members of the Society.

Two Council meetings were held during the Conference on the evenings of August 29 and September 1. They were attended by Miss Cox and Messrs. Hopkins, Mahalanobis, Schwartz, Neurdenburg, Fisher, Cochran, Bliss, Buzzati-Traverso, Linder, Finney and Yates. The meetings were concerned primarily with problems arising in the Conference or which will appear in later issues of *BIOMETRICS*.

All members of the International Statistical Institute, with which we are affiliated, were invited to participate in the 2nd International Biometric Conference. Similarly, all those attending the Conference in Geneva were invited to attend the meetings of the International Statistical Institute in Berne during the following week. Many of those attending both conferences took advantage of a special train which left Geneva at 8:30 AM September 3 for Berne via Sion, Lotschberg, Interlaken, Thun and Lucerne. The weather conditions were perfect and all enjoyed some magnificent views of Swiss mountain scenery.

The success of the Conference was due in large part to the excellent work of Professor Arthur Linder who served as Secretary of the Conference Committee. Those of us who had the good fortune to attend the meetings will long remember the many courtesies of Professor Linder and his aides.

ISI meetings in Berne. The 26th Session of the International Statistical Institute convened at the University in Berne on September 5. Fifteen of the papers on the ISI program were by members of the Biometric Society and lay in the fields of statistical sampling, industrial applications of statistical methods, statistical education, recent developments in statistics and demography. These sessions enjoyed the

same good weather as those in Geneva. One meeting of the Bureau of the ISI was attended by representatives of affiliated organizations including the Biometric Society, and considered how we can obtain the greatest advantage from our affiliations.

An International Statistical Seminar under the auspices of the ISI was held during the two weeks following the meetings at Berne. The first week at the University in Berne on September 12-17 included several lectures on experimental design and industrial applications by members of the Biometric Society including Professors Cox, Bliss, Quenouille, Linder and Day. For the second week the sessions were shifted to Geneva and emphasized statistical sampling with Professors Darmonis, Deming, Linden, Madhava, Mahalanobis and Yates among the lecturers.

Project on Training in Biometry. The session on teaching and education of biometry at Geneva revealed the need for a wide exchange of information on the material covered by courses in this field. Syllabi are wanted showing the time spent on each topic in courses on biometry or including biometry in different universities of the world. The Society has been asked to assemble such information and make it available to teachers in the field. The subject is also of interest to UNESCO and certain of its affiliated agencies. Plans are now being made to assemble this information and progress will be announced in later issues of BIOMETRICS. A committee consisting of Professors W. G. Cochran (Chairman), C. I. Bliss, A. Buzzati-Traverso, G. Darmonis and K. Mather has been named by President Fisher to undertake this project.

NEWS AND NOTES

BIOMETRIC SECTION OF THE AMERICAN STATISTICAL ASSOCIATION ANNUAL MEETING, DECEMBER 28-30, BILTMORE HOTEL, NEW YORK CITY: Sessions arranged by the Biometric Section. Joining organization: Biometric Society.

Wednesday, December 28, 4-6.

Topic: The use of rationally developed equations in biology. Chairman: Horace W. Norton.

Papers: An interpretation of the formation of active bacterial virus from ultraviolet inactivated virus. S. E. Luria. The application of equations derived from models, to "central" circulatory volume. Elliot V. Newman and Margaret Merrell.

Discussants: Joseph Berkson and L. J. Savage.

Thursday, December 29, 2-4.

Topic: Long-time follow-up in morbidity studies. Chairman: John W. Fertig.

Papers: The definition of the group to be followed. Paul M. Densen. Timing of the distribution of the events between observations. T. E. Harris, Paul Meier and John W. Tukey. Methods of analysis in follow-up studies. Harold F. Dorn.

Discussants: Hugo Muench, Rowland Rider and Mortimer Spiegelman.

Friday, December 30, 10-12.

Topic: Contributed papers. Chairman: Frederick Mosteller.

Papers: Relative precision of minimum χ^2 and maximum likelihood estimates of regression coefficients, with particular reference to bioassay. Joseph Berkson. Malformations at the Boston Lying-in Hospital, 1930 to 1941. Jane Worcester and Stuart S. Stevenson. A statistic for rating diagnostic tests. W. J. Youden.

Discussants: Paul Bruyere, Chester I. Bliss and John Tukey.

NORTH CAROLINA INSTITUTE OF STATISTICS TASTE TESTING CONFERENCE—On November 7-10, the University of North Carolina Institute of Statistics held another in a series of statistical work conferences under the sponsorship of the General Education Board, this time in the field of taste testing. About 25 leaders in the field were present. The program included various aspects of organizing and con-

ducting taste tests with particular emphasis upon statistical procedures available for this type of experimentation. Some of the subjects discussed were: "Fundamentals of Flavor Characterization," by E. C. Crocker of Arthur D. Little, Inc., Cambridge, Massachusetts; "Layout and Design of Flavor-Preference Panels," by W. Franklin Dove, Food Consultant, Oak Park, Illinois; "Scoring Systems," by J. W. Hopkins, Division of Applied Biology, National Research Laboratories, Ottawa, Ontario, Canada; techniques of laboratory testing; food surveys; the replacement of organoleptic tests by physical-chemical methods; and designs which are appropriate for tasting experiment.

AUSTRIA—The following quotation was taken from a letter from W. Winkler, Director of the Institute of Statistics of the University of Vienna with whom we have recently agreed to exchange *Biometrics* for *Statistische Vierteljahresschrift*. "Our Institute of Statistics corresponds rather to the Department of Statistics in your universities or still to less. The hitherto regulations about statistical teaching are rather poor, but on the way to being replaced by more sufficient ones. Also the introduction of courses for 'professional statisticians' is in preparation with the aim of getting the title of a 'diploma statistician'. As all that is only on the way and the present state of things poor, I renounce describing it and request you kindly to have patience till I am in a position to write of reforms already performed."

CZECHOSLOVAKIA—From the School of Agriculture and Forestry, Czech Technical University, Praha, Czechoslovakia, Vaclav Myslivec writes, "At our Czech Technical University, faculty of agriculture and forestry, I am giving a couple of courses. These courses are: Elements of higher mathematics, Biometrics and Experimental Statistics. The second course is given for the students of forestry and the third for students of agriculture. In the second course I am giving fundamentals of sampling theory, which is very important for forestry mensuration and taxation. In the third course, which is highly important for agricultural experimentation, I am giving lectures on fundamentals of mathematical statistics, and most of the time is given to analyses of variance and to design of laboratory and field experiments. I have found a lot of interest among my students and particularly among research workers not only in the school but also in the experimental stations. The first course, concerning calculus, is a preparatory course for my second and third course. I really enjoy my work in this field of science. The main reason for my joy is that I found a rather big audience and secondly that I am bringing to my pupils quite a new and important knowledge.

This knowledge is of fundamental importance for all research work in agriculture and forestry."

HAWAII—O. E. Sette of the United States Fish and Wildlife Service informs us that his recent move from San Francisco to Honolulu was in connection with the establishment of a new activity within the Fish and Wildlife service to be known as the Pacific Oceanic Fishery Investigations of which Mr. Sette is Director. **Milner Baily Schaefer**, formerly with the South Pacific Investigations of the United States Fish and Wildlife Service with headquarters at Stanford University, California, has been transferred to the Pacific Oceanic Fishery Investigations. The staff is now occupying temporary quarters furnished by the Navy but expects to be able, sometime after the first of January, 1950, to move into a new laboratory being constructed on the campus of the University of Hawaii. These Investigations are engaged in studies toward the development of the now unutilized high seas fisheries of the Pacific oceanic areas embracing the region to the southward of the Hawaiian Islands, and the areas to the westward which were formerly under a Japanese mandate but now constitute the Trust Territories of the Pacific. The Section of Biology and Oceanography of which Mr. Schaefer is chief, is engaged in studies of the biology, ecology and distribution of the tunas and other pelagic fishes and the relationship thereof to the various factors of the oceanic environment such as currents, temperature distributions, and other physical and chemical factors.

UNITED STATES—For having written and published an article entitled "Casualties of the United States Eighth Air Force in World War II", **James A. Rafferty**, chief of the Department of Biometrics at the U.S. Air Force School of Aviation Medicine, Randolph Air Force Base, has received commendations from **Brig. Gen. Otis O. Benson, Jr.**, Commandant of the School of Aviation Medicine, and **Col. George F. Baier III**, surgeon of the Air University at Maxwell Air Force Base, Ala. The report was startling in that it revealed that about one half of the casualties reported were of non-battle types. . . . **Fred A. Schultz**, Director of Pharmaceutical Research, Commercial Solvents Corporation, Terre Haute, Indiana, informs us that his interest is concerned with the statistical evaluation of data obtained in animal experimentation. "At the present time we are carrying out a large number of experiments for the evaluation of compounds for their possible use as therapeutic agents. Needless to say, in the evaluation of our animal results a statistical analysis of the data is essential." . . . At Iowa State, **Gerhard Tintner** has returned to the staff of the Statistical Laboratory and the Economics

Department after a year's leave of absence to work at Cambridge University in the Department of Applied Economics; **Robert G. D. Steel** has taken a position with the College of Agriculture at the University of Wisconsin, after receiving his Ph.D. at Ames in June. Mr. Steel's dissertation was, "Minimum Generalized Variance for a Group of Linear Functions"; **Bernard Ostle** received his Ph.D. in Statistics at the end of the summer session with a dissertation "On Certain Criteria for Optimum Estimation." He remains at Iowa State as Assistant Professor in the Department of Statistics; **Osmer Carpenter** received his Ph.D. in June and returned to his position with the Atomic Energy Commission at Oak Ridge, Tennessee. His dissertation was, "Sequential Tests of the Linear Hypothesis"; **Douglas Robson** obtained his B.S. in Statistics in June and is now working with **Walter Federer** in the Department of Plant Breeding at Cornell University; joining the ranks of the thirty-odd graduate students majoring in Statistics in the Department are **Om Prakash Aggarwal** from Delhi University, India, and **Bertil Matérn**, a student of Cramér, from the Forest Research Institute, Sweden. . . . **E. L. Cox** formerly at Virginia Polytechnic Institute and who has just spent a summer in Eastern Canada chasing fishes for the Atlantic Salmon Investigation, is now with the Institute of Statistics working on the application of statistics to fishery research problems. He is, to quote Mr. Cox, "figuring how to catch more fishes by putting statistics instead of salt on their tails." Also with the Institute are **Dan Teicherow**, from the University of Toronto who is working on a doctorate in Experimental Statistics, and **A. Grandage** from Schering Corporation, Bloomfield, New Jersey, whose major interest is in the statistical aspects of biological assay. . . . From the Department of Mathematics, Statistical Laboratory at the University of California, Berkeley, we have a report on recent changes of status. **Henry B. Mann** of Ohio State University has accepted a Visiting Professorship and Research Associateship for the academic year 1949-1950; **J. Neyman**, Director, will be on sabbatical leave for the Spring Semester, 1950; **Joseph L. Hodges, Jr.** has been promoted to Assistant Professor and Research Associate; **Charles M. Stein**, Assistant Professor and Research Associate, will be on leave for the academic year 1949-1950, and will be working in Paris as a National Research Fellow; **Douglas G. Chapman**, **Mark W. Eudey**, **Elizabeth L. Scott** and **Ester Seiden** obtained their Ph.D. degrees in Statistics at the University of California. Douglas G. Chapman has accepted an Assistant Professorship at the University of Washington, Seattle. Mark W. Eudey is now Vice-President of California Municipal Statistics. Elizabeth L. Scott and Ester Seiden have been promoted to Lecturer and Research Associate at the Statistical Laboratory.

BIOMETRICS

VOLUME 5: 1949

Numbers 1-4

Published Quarterly by the
AMERICAN STATISTICAL ASSOCIATION
WASHINGTON, D. C.
1949

STATEMENT OF THE OWNERSHIP, MANAGEMENT, AND CIRCULATION REQUIRED BY THE ACT OF CONGRESS OF AUGUST 24, 1912, AS AMENDED BY THE ACTS OF MARCH 3, 1933, AND JULY 2, 1946 (Title 39 U. S. C. 233) of *Biometrics*, published quarterly at Washington, D. C. for 12 months ending October 1, 1949.

1. The names and addresses of the publisher, editor, managing editor, and business managers are: Publisher, American Statistical Association, 1603 K St., N. W., Washington 6, D. C.; Editor, Gertrude M. Cox, Institute of Statistics, Raleigh, N. C.; Managing editor, None; Business manager, None.

2. The owner is: (If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as those of each individual member, must be given.) American Statistical Association, 1603 K St., N. W., Washington 6, D. C.

3. The known bondholders, mortgages, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities are: (If there are none, so state.) None.

4. Paragraphs 2 and 3 include, in cases where the stockholder or security holder appears upon the books of the company as trustee or in any other fiduciary relation, the name of the person or corporation for whom such trustee is acting; also the statements in the two paragraphs show the affiant's full knowledge and belief as to the circumstances and conditions under which stockholders and security holders who do not appear upon the books of the company as trustees, hold stock and securities in a capacity other than that of a bona fide owner.

5. The average number of copies of each issue of this publication sold or distributed, through the mails or otherwise, to paid subscribers during the 12 months preceding the date shown above was: (This information is required from daily, weekly, semiweekly, and triweekly newspapers only.)

SAMUEL WEISS, *Executive Director*

Sworn to and subscribed before me this 14th day of October, 1949.
(My commission expires June 1, 1953.)

MARGUERITE CONNORS, *Notary Public*

BIOMETRICS

Contents of Volume V

ABSTRACTS	76
NEWS AND NOTES	94, 183, 256, 345
QUERIES	68, 174, 250, 335
THE BIOMETRIC SOCIETY	90, 180, 254, 341
 <i>ARTICLES and ABSTRACTS</i>	
ANSCOMBE, F. J. The Statistical Analysis of Insect Counts Based on the Negative Binomial Distribution	165
BALDWIN, A. L. <i>Abstract</i> #59	76
BARTLETT, M. S., Fitting a Straight Line when Both Variables are Subject to Error	207
BERKSON, JOSEPH, <i>Abstract</i> #79,	86
#81	87
BERNSTEIN, M. E., <i>Abstract</i> #80	86
BLISS, C. I., and NEEELY TURNER, <i>Abstract</i> #72	83
BROWN, G. W., <i>Abstract</i> #67	80
CHASSAN, J. B., On a Statistical Approximation to the Infection Interval	243
COLLINS, S. W., <i>Abstract</i> #70	82
CORNFIELD, JEROME and NATHAN MANTEL, <i>Abstract</i> #77	85
CRONBACH, L. J., <i>Abstract</i> #63	78
FEDERER, W. T. The General Theory of Prime-Power Lattice Designs. III. The Analysis for p^3 Varieties in Blocks of p Plots with more than 3 Replicates	144
FINNEY, D. J. The Choice of A Response Metameter in Bio-Assay	261
FISHER, R. A. A Biological Assay of Tuberculins	300
FRY, F. E. J. Statistics of a Lake Trout Fishery	27
GARDNER, ERIC F. <i>Abstract</i> #68	80
GREEN, M. W. <i>Abstract</i> #73	83
GUETZKOW, HAROLD <i>Abstract</i> #66	79
HARSHBARGER, BOYD. Triple Rectangular Lattices	1
HARSHBARGER, BOYD, <i>Abstract</i> #78	85
HEALY, M. J. R. Routine Computation of Biological Assays Involving a Quantitative Response	330
HORN, DANIEL, <i>Abstract</i> #65	79
HUNT, J. M., see LEONARDS Kogan	
IPSEN, J., <i>Abstract</i> #74	84
JERNE, N. K. and E. C. WOOD, The Validity and Meaning of the Results of Biological Assays	273
KOGAN, LEONARDS and J. M. HUNT, <i>Abstract</i> #61	77
KUBIS, J. F., <i>Abstract</i> #60	76
MANTEL, NATHAN, see JEROME CORNFIELD	
MARCUSE, SOPHIE. Optimum Allocation and Variance Components in Nested Sampling with an Application to Chemical Analysis	189
MASUYAMA, MOTOSABURO. On a One-Dimensional Diffusion Method of Assaying Antibiotic Substances and Its Fundamental Formulas	317
MATHER, K. The Analysis of Extinction Time Data in Bio-assay	127
MILLER, R. B. Problems of the Optimum Catch in Small Whitefish Lakes	14
QUENOUILLE, M. H. A Relation Between the Logarithmic, Poisson and Negative Binomial Series	162
RABIN, A. I., <i>Abstract</i> #62	77
REIERSOL, OLAV, <i>Abstract</i> #82	88
ROUNSEFELL, G. A. Methods of Estimating Total Runs and Escapements of Salmon	115

SANDERS, B. S., <i>Abstract</i> #71	82	vidual Means in the Analysis of Variance	99
SCHUCK, H. A. Relationships of Catch to Changes in Population Size of New England Haddock	213	One degree of Freedom for Non-Additivity	232
STEPHENSON, WILLIAM, <i>Abstract</i> #64	78	TURNER, NEELY, see C. I. BLISS	
TUKEY, J. W., <i>Abstract</i> #75	84	VOTAW, D. F., JR., <i>Abstract</i> #69	81
#76	84	WOOD, E. C., see N. K. JERNE	
TUKEY, J. W. Comparing Indi-			

INDEX

- Additivity, 84, 232
Adjusted means, 11, 157
Age distribution, 17, 31, 219
Agronomy, 252
Analysis of covariance, 72, 178, 289
Analysis of variance, 4, 72, 84, 85, 99, 149, 176, 192, 232, 250
Animal breeding, 71
Approximate significance levels, 102
Bacteriology, 127, 162, 317, 333
Binomial series, 162, 165
Biological assay, 83, 86, 127, 211, 261, 273, 300, 317, 330
Biometric Society, 90, 180, 254, 341
Biometry, 86, 243
Case histories, 76
Chemistry, 189
Chi square, 83, 86, 140, 251, 301
Clinical statistics, 76
Coefficient of variation, 161
Components of variance, 74, 155, 189
Computing methods, 330
Confidence limits, 178, 212
Confounding, 74, 145, 158
Correlation, 80, 42, 123
Curve fitting, 207
Cycles, 120, 125
Degrees of freedom, 232
Dependent variable, 73
Design of experiments, 1, 70, 85, 144, 169, 174, 178, 278, 283, 289
Discriminant functions, 80
Distributions, 109, 162, 165
Dosage-mortality curves, 83, 85
Drug dosages, 83
Efficiency, 68, 156, 167, 208, 288
Empirical formula, 319
Entomology, 68, 83, 165
Epidemiology, 243
Error mean square, 72, 178
Error theory, 301
Estimation, 68, 84, 131, 166, 168, 192, 208, 288, 290
Expectations, 308
Experimental sampling, 111
Extinction time, 127
Extreme deviate, 107, 250
Fiducial limits, 177, 266, 288, 290, 311
Field experiments, 1, 85, 144, 252
Fisheries, 14, 27, 115, 213
Fitting curve, 207
Fitting distribution, 165
Fractional replication, 73
F-test, 241
Growth, 42
Incomplete blocks, 144, 154
Identification, 88
Independent variable, 73
Infection interval, 243
Information, 155
Judgements, 77
Largest difference, 99, 250
Latin square, 1, 71, 102, 170, 305
Lattice, 1, 144
Least squares, 208, 212
Linear hypothesis, 73
Logarithmic series, 162
Logarithmic transformation, 72
Loglog transformation, 130
Mathematical biology, 27, 117, 243, 323
Mathematical model, 73
Maximum likelihood, 83, 85, 86, 130
Medical statistics, 95
Morbidity, 82, 243
Mortality, 24, 34, 57
Multivariate analysis, 81
Negative binomial, 162, 165
Nested sampling, 189
Nomography, 330
Normal distribution, 325
truncated, 84
Normality, 263, 279, 282, 287
departure from, 265
Null hypothesis, 280
Nutrition, 178, 250
Optimum catch, 14, 24
Optimum sampling, 189
Physiology, 176

- Poisson distribution, 83, 129, 162
Population management, 14, 125
Pseudo-effects, 145
Pseudofactors, 145
Psychology, 76, 77, 78, 79, 80
Quadrat sampling, 163
Quality control, 83
Randomization, 178, 279, 286
Randomized blocks, 154, 174
Regression, 68, 87, 88, 123, 140, 228, 237
 multiple, 119
 partial, 175
Replication, 73, 179
Reproduction, 49
Sampling, 111, 174, 189, 205
Scatter diagram, 30, 41, 51, 122, 230, 238, 247
Scedasticity, 261, 283, 287
Sex ratio, 86
Significance, 102, 236
Skewness, 263
Soil science, 174
Standard error, 11, 141, 159
Statistical model, 261, 273
Stratified sampling, 174, 205
Studentized deviate, 107
Surveys, 82, 165
Teaching statistics, 94, 96
Transformations, 73, 86, 130, 166, 167, 172, 241, 261, 283
Trend, 41
Triple lattice, 1
t-test, 101, 124
Tuberculin, 300
Uniformity trials,[†]144
Variance, 6
 allocation of, 189
 interblock, 5, 151
 intra-block, 5, 154
 sampling, 140
 see also Analysis of—, Components of—
Weights, 5, 306
Yield trials, 252

**Indian Agricultural Research
Institute Library, New Delhi**

1. Books are issued for 14 days, beyond if for next 7 days a fine of 15 paise & after wards 25 paise per day shall be charged.
2. Books may be renewed at the discretion of the head Library Services.
3. Dog-earing of pages of a book, marking of writing there in with ink or pencil, tearing or taking out its pages or otherwise damaging it will constitute an injury to a book.
4. Unless a borrower points out the injury at the time of borrowing the book, he shall be required to replace the book or pay its price, If detected at the time of return.

**HELP US TO KEEP THIS BOOK
FRESH AND CLEAN**

